DROUGHT OVER THE PAST CENTURY IN TEXAS AND NEW MEXICO: REDUCING INHOMOGENEITIES IN LONG-TERM CLIMATE RECORDS VIA STATISTICAL METHODS TO STUDY DROUGHT

A Thesis

by

DOUGLAS BRENT MCROBERTS

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

May 2008

Major Subject: Atmospheric Sciences

DROUGHT OVER THE PAST CENTURY IN TEXAS AND NEW MEXICO:

REDUCING INHOMOGENEITIES IN LONG-TERM CLIMATE RECORDS VIA

STATISTICAL METHODS TO STUDY DROUGHT

A Thesis

by

DOUGLAS BRENT MCROBERTS

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Approved by:

Chair of Committee,	John Nielsen-Gammon
Committee Members,	Fuqing Zhang
	Steven Quiring
Head of Department,	Kenneth P. Bowman

May 2008

Major Subject: Atmospheric Sciences

ABSTRACT

Drought over the Past Century in Texas and New Mexico: Reducing Inhomogeneities in Long-term Climate Records via Statistical Methods to Study Drought. (May 2008)

Douglas Brent McRoberts, B.S., Purdue University

Chair of Advisory Committee: Dr. John Nielsen-Gammon

This research looks at the past century of Texas and New Mexico climate in order to create datasets sufficient for documenting climatic variations. Inhomogeneities in climate records are defined as variations in climatic records caused by factors other than weather and climate. While there are indirect methodologies for inferring climate records such as tree rings and ice cores, it is the instrumental network that constitutes the most spatially and temporally complete record of land surface climate since the onset of the Industrial Revolution. A statistical method by Sun and Peterson (2005a) called Inverse Weighting of Square Distance (IWSD) will be used to reduce the inhomogeneities in climate records.

The National Weather Service Cooperative Observer Program (COOP) network of stations will be used for this analysis. A subset of the extensive COOP network, called the United States Historical Climate Network (USHCN), will be used as a foundation for this study. The analysis and resulting datasets from this climatic study show precipitation trends and periods of drought and will be useful for decisions regarding future policies on drought.

The result of the interpolation process was the creation of several COOP and USHCN datasets. Several of the datasets were investigated to determine the spatial characteristics of precipitation over the 20th century in Texas and New Mexico. The datasets are in good agreement that the most severe drought period of the 20th century in Texas and New Mexico was in the 1950s. The frequency of pluvial periods was higher toward the end of the 20th century, with most USHCN stations showing an increasing trend when a linear regression analysis was done on each station's precipitation data.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Nielsen-Gammon, and my committee members, Dr. Zhang, and Dr. Quiring for their guidance and support throughout the course of this research.

Thanks also go to my friends and colleagues and the department faculty and staff for making my time at Texas A&M University a great experience. I also want to extend my gratitude to the National Climate Data Center, which provided the data that was the foundation of my research.

Finally, I would like to give thanks to my mother and father. Though they are a thousand miles away, they have supported me through all my ups and down during my time at Texas A&M University.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE Page

FIGURE Page

FIGURE Page

FIGURE Page

LIST OF TABLES

TABLE Page

1. INTRODUCTION

An understanding and clarification of past climatic trends can help future predictions of drought and other extreme events. In accordance with Task 1 of a project created to look at "Change, drought, and policy making in the United States Southern Region", the goal of this research is to study drought in New Mexico and Texas. Climate data will be analyzed on a number of time scales to identify periods of drought and pluvial conditions over the past century. Drought conditions are periods of time characterized by a lack of precipitation while pluvial conditions are times marked by excess precipitation.

While there are indirect methodologies for inferring climate records such as tree rings and ice cores, it is the instrumental network that constitutes the most spatially and temporally complete record of land surface climate (Jones 1995). Most instrumental networks were established to monitor local weather and not the long-term climate; there are practical problems in using these data to study climate change. Therefore, it is rare to find stations with even relatively homogeneous time series whose fluctuations are consistent with those from surrounding stations (Conrad and Pollack 1962).

The foundation of long-term climate studies in the United States is the National Weather Service Cooperative Observer Program (COOP) network (NCDC 2006), which includes several hundred stations in Texas, New Mexico, and surrounding states with various periods of rainfall and temperature records. The COOP was formally created in 1890 under the Organic Act, with eventually more than 11,000 volunteers taking observations on farms, in urban and suburban areas, national parks, seashores, and

---

This thesis follows the style of *Journal of Climate*.

mountaintops.

Currently there are approximately 8,000 active stations in the COOP data set. Historically, approximately 23,000 stations have recorded data at some point during the 20th century (NCDC 2006). Unfortunately, most long-term climatological time series have been affected by a number of non-climatic factors that make the raw data unrepresentative of the actual climate variation occurring over time. These factors include changes in: instruments, observing practices, station locations, formulas used to calculate means, and station environment (Peterson et al. 1998). Because the network consists of mostly volunteers, there are numerous gaps in the COOP record and stations come and go in irregular patterns.

A subset of the COOP data is the United States Historical Climate Network (USHCN) data set, consisting of precipitation and temperature records from 1,221 stations spanning most of the 20th century. The USHCN is a high-quality long-term data set of monthly averaged maximum, minimum, and mean temperature and total monthly precipitation developed to assist in the detection of regional climate change (Karl et al. 1990).

The USHCN was developed using stations with very few gaps in their records, but the spatial coverage of the network is poor. Detection of more localized climatic signals is only possible through indirect methodologies that infer missing data by means of statistical approaches. The USHCN project dates back to the middle 1980s and was created in response to the need for an accurate, unbiased, modern historical climate record for the United States. Personnel at the Global Change Research Program of the Department of Energy and at NCDC defined a network of 1,221 stations in the contiguous United States whose observations would comprise a key baseline dataset for monitoring climate (Karl et al. 1990). Since then, the USHCN dataset has been updated several times, the most recent of which includes data through 2005. However, because the availability of this dataset was too late to be included in this study, the 2001 version was used.

A homogeneous climate time series is defined as one whose variations are only caused by variations in weather and climate (Peterson et al. 1998). A significant variation of a single station climate signal from a regional climate signal may be due to inhomogeneities. Inhomogeneities can be manifested in a climatic time series in one of two ways: as a sharp discontinuity or as a gradual trend away from a regional climate signal (Peterson et al. 1998). The first is indicative of an abrupt change such as a shift in station location, while the second indicates a slow deterioration of instrumentation or changes in land use surrounding a station. The fundamental problem with most climate studies is that long-term trends do not take sharp discontinuities into account when making calculations (National Research Council 2002).

By examining a particular station's climate record by using climate records from surrounding stations, the more significant inhomogeneities can be identified through statistical methods. The statistical methodology that best fits the data and the needs of our project is that documented by Sun and Peterson (2005a). The technique used by Sun and Peterson can be applied to both precipitation and temperature data and creates a value for a target station based on data from several surrounding stations.

The Sun and Peterson (2005a) technique can also utilize all the information from an incomplete climate record for a particular COOP station to fill in missing data. This is done by forming statistical relationships between the available COOP data and the USHCN data. Therefore, in periods when data for a particular station is incomplete, we can indirectly infer the monthly precipitation or temperature value based on a modified Sun and Peterson (2005a) technique.

The purpose of this study is to look for climate variability affecting drought and pluvial conditions in New Mexico and Texas. Several of the COOP stations have short periods of record that are not useful for longer term climatic studies. These missing data can be inferred with the statistical interpolation, though a degree of uncertainty will exist in this attempt to recreate missing data.

Inferring station history using surrounding stations will create a "relative" homogeneity to the climate records. Precipitation measurements are particularly susceptible to large inhomogeneities and biases (Easterling et al. 1996). Precipitation can vary significantly across a relatively small distance, particularly for individual events. Our study will focus mainly on monthly precipitation totals, a quantity more easily derived using surrounding station data than daily totals.

However, interpolation will still be useful in unearthing smaller scale spatial variations in precipitation records. By nature, temperature has fewer local variations, but the number of available COOP stations reporting daily maximum and minimum temperature data is significantly less than those reporting daily precipitation totals. The number of available stations varies with time, reaching a peak in the 1950s. Figure 1 shows the number of COOP stations with data through time for the entire state of Texas.

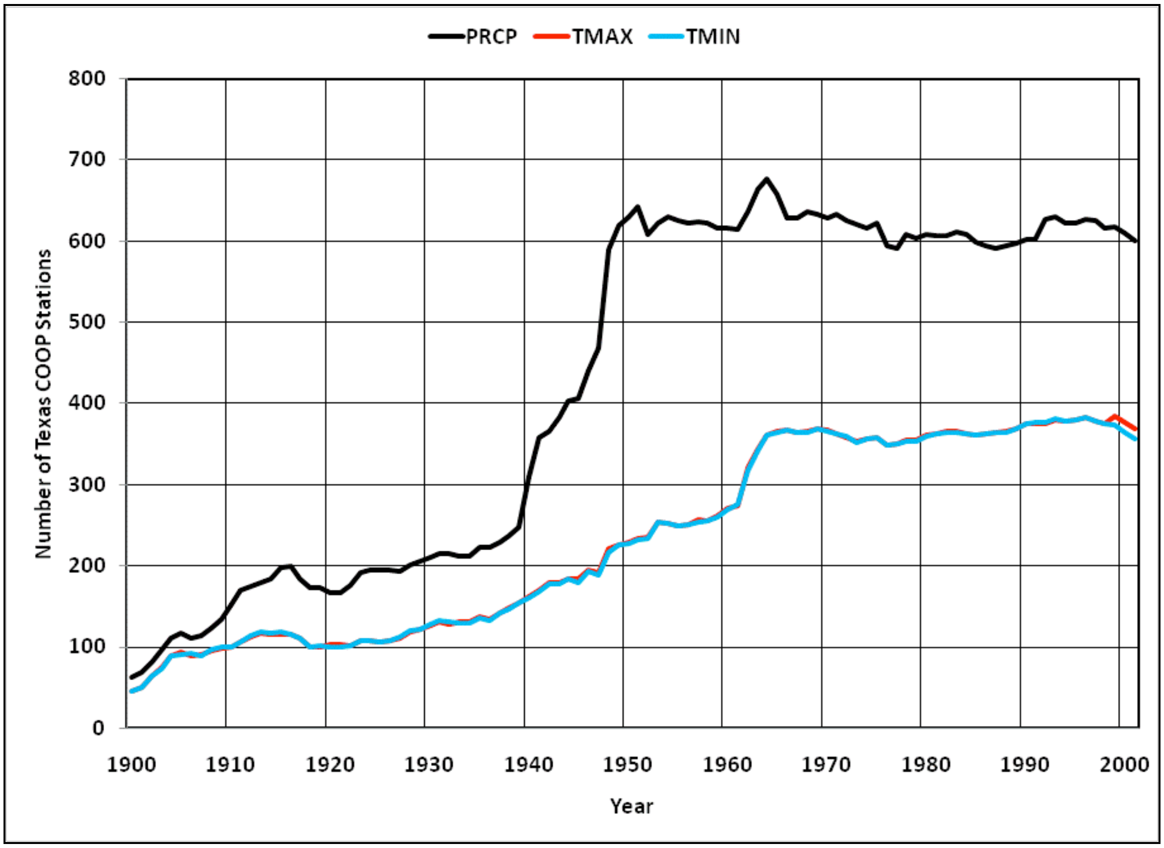


FIG 1.1. The total number of COOP stations in the state of Texas as a function of time. Specifically shown are precipitation (black), maximum temperature (red), and minimum temperature (blue).

The most noticeable discontinuity in the graph is the sudden increase around 1950 of stations reporting precipitation. Beginning in 1948 (when surplus keypunch machines were obtained from the U.S. Postal Service by the then-existing New Orleans branch of NCDC, cooperative observations were routinely stored on machine-readable punch cards (Kunkel et al. 2005). The number of temperature measurements exhibits a roughly linear increase with a discontinuity at the same time as the precipitation discontinuity, but with a smaller magnitude.

The end result of the data processing and interpolation will be a more spatially complete dataset of long-term trends in precipitation. Since this is not a study in the quality-control of the data, no additional modifications are performed on the two data sets from NCDC other than the elimination of outliers for which an extensive procedure will be described. We will form a network similar to the USHCN for Texas but with a significant increase in the number of stations so that recent trends in climate are more spatially resolved. This will allow for a better resolution of patterns in trends that may not be well resolved with the current network of stations we have in place.

Once the virtual data network has been derived from the statistical methods, analysis of the resulting data will look at spatial patterns and temporal trends in the precipitation records and relate these to drought. Droughts over the past several years have had important implications on water use throughout Texas. The network data created by this study will be used by the Spatial Sciences laboratory to create a high resolution climate atlas of Texas.

However, it is important to note that only precipitation data will be analyzed in this study. The procedures describing the quality control and interpolation of maximum and minimum temperature will be described, but the analyses of the datasets will deal *entirely* with precipitation.

2. LITERATURE REVIEW ON DROUGHT

a. Definition

Drought is a commonly used term used to describe periods on different temporal scales describing a general lack of precipitation. The general concepts used today as meteorological definitions on dry periods are consecutive days with no precipitation, consecutive days with little precipitation, or little precipitation during a specific period of time (Byun and Wilhite 1999). During the first decade of the twentieth century, the U.S. Weather Bureau identified drought as occurring during any period of 21 or more days with rainfall 30% or more below normal for the period (Henry 1931). Friedman (1957) used annual rainfall as his drought index in a study of drought in Texas.

Drought affects people with varying interests in so may different ways that it is impossible to create a unique definition. The problems with developing an agricultural drought index, for example, include consideration of vegetation, soil type, antecedent soil moisture, and evapotranspiration as influenced by wind speed and the temperature and humidity of the air (Heim Jr. 2002).

Drought has been defined by the international meteorological community in several ways (Heim Jr. 2002). Because drought implies different things to different people, there is no consistent definition for drought (Mo and Chelliah 2006). The confusion over drought is that it is an intricate concept that cannot always be completely characterized by a number or by statistics. Some locations have more variability in their precipitation climate than other locations, relying on a small number of heavy precipitation events while other locations rely on more frequent moderate precipitation events.

For example if there is month in a given location in which five times more precipitation than the mean falls but is between two months in which little or no precipitation fell, how is this characterized? Most current indices assess only the deficiency of water from the climatological mean for some predefined duration (Byun and Wilhite 1999). Also, there is no consistent indicator for the conclusion of a drought period. Can one month of extreme pluvality overcome several months of dry conditions or does it take several months of at least normal precipitation conditions?

b. Palmer Drought Severity Index (PDSI)

Several different drought indices have been created throughout the years to deal with drought based on specific interests or groups of people. Palmer (1965) created the Palmer Drought Severity Index (PDSI) that incorporated antecedent precipitation, moisture supply, and moisture demand into a hydrologic accounting system. Of all the indices, the PDSI is still the most widely used and recognized index on an operational basis (Byun and Wilhite 1999).

The objective of the PDSI is to provide measurements of moisture conditions that are standardized so that comparisons using the index can be made between locations and between months (Palmer 1965). The usefulness of the PDSI is that it has a single Z-value describing the existence of drought, pluvial, or neutral conditions at a given location. The following table (Table 2.1) shows the definitions for numerical groupings of the Z-values resulting from the PDSI algorithm.

TABLE 2.1. Palmer Drought Severity Index categories.

PDSI Value	Conditions
PDSI ≥ 4.00	Extremely wet
3.00 ≤ PDSI ≤ 3.99	Very wet
2.00 ≤ PDSI ≤ 2.99	Moderately wet
1.00 ≤ PDSI ≤ 1.99	Slightly wet
0.50 ≤ PDSI ≤ 0.99	Incipient wet spell
0.49 ≤ PDSI ≤ -0.49	Near normal
-0.99 ≤ PDSI ≤ -0.50	Incipient dry spell
-1.99 ≤ PDSI ≤ -1.00	Mild drought
-2.99 ≤ PDSI ≤ -2.00	Moderate drought
-3.99 ≤ PDSI ≤ -3.00	Severe drought
PDSI ≤ -4.00	Extreme drought

Ideally, the PDSI is designed so that a -4.0 in South Carolina has the same meaning in terms of the moisture departure from a climatological normal as a -4.0 in Idaho (Alley 1984). Wells et al. (2004) however argues that the behavior of the PDSI at various locations is inconsistent, making spatial comparisons of PDSI values difficult, if not meaningless. Because of the complexity of drought, no single index has been able to adequately capture drought and its potential impacts on a diverse population (Heim Jr., 2002).

c. Standardized Precipitation Index (SPI)

Because of the intricacies involved in calculating the PDSI, it was to find a simpler precipitation-based method to study drought and pluvial conditions. The Standardized Precipitation Index (SPI) calculation for any location is based on the long-term precipitation record for a desired period. This long-term record is fitted to a probability distribution so that the mean SPI for the location and desired period is zero

(Edwards and McKee 1997). Table 2.2 describes the classifications of conditions in a given location based on its SPI value for a given time.

TABLE 2.2. Standardized Precipitation Index categories.

SPI Value	**Conditions**
SPI $\geq$ 2.00	Extremely wet
1.50 $\leq$ SPI $\leq$ 1.99	Very wet
1.00 $\leq$ SPI $\leq$ 1.49	Moderately wet
-0.49 $\leq$ SPI $\leq$ 0.49	Near normal
-0.49 $\leq$ SPI $\leq$ -1.00	Moderately dry
-1.99 $\leq$ SPI $\leq$ -1.50	Severely dry
SPI $\leq$ -2.00	Extremely dry

The SPI method of fitting a probability distribution to a time series of precipitation values is desirable for the precipitation time series datasets at the disposal of this study. Historical data are used to compute the probability distribution of the monthly and seasonal observed precipitation totals, so that the SPI values can be calculated for different time scales. Analyses of drought based on precipitation will be a modification of the SPI methodology and discussed in further detail in later sections.

d. Other Drought Indices

The numerous other drought indices created are often designed as the modification of an existing scheme. Wells et al. (2004) modified the PDSI to the SC-PDSI (Self Calibrating PDSI) that continuously modifies the empirical values in the formulas to fall in line with the climate expectancies at a given location. Mo and Chelliah (2006) also modify the PDSI using the National Centers for Environmental Prediction (NCEP) North American Regional Reanalysis (RR) from 1979 to 2004. Many deficiencies of the original PDSI are eliminated by taking fields directly from the RR or by making better estimates. Katz and Glantz (1986) use a "Standard Anomaly Index" that seeks capture the spatial characteristics of drought in a single number.

e. History of Drought Conditions in Texas and New Mexico

Several studies indicate that droughts in the 1930s and 1950s were among the worst in the United States. Andreadis et al. (2005) looked at drought using a series of severity-area-duration (SAD) curves and determined the droughts in these two decades were the most severe to affect large areas. The drought of the 1930s was associated with the Dust Bowl era and affected mainly the northern areas in our domain. In fact, East Texas was one of the few locations in the entire United States to escape drought conditions at the height of the Dust Bowl era (Hecht 1983).

The reanalysis of PDSI values done by Hecht (1983) showed, however, that nearly all of New Mexico had PDSI value below -4.0 at the height of the 1950s drought, which Table 2.1 shows to be an extreme drought. Another finding of Hecht (1983) was that the pluvial conditions of Texas in April 1977 with PDSI values above 2.0 for most of the state coincided with an extreme drought in the northern tier of the United States.

Stahle and Cleaveland (1988) took detailed look specifically at Texas drought and pluvial conditions from 1680-1980, with particular interest in the month of June as an indicator. In this time period Stahle and Cleaveland (1988) reconstructed PDSI values and state that the driest years in the 20th century were 1925, 1971, 1917, and 1956, while 1919 and 1924 are the wettest 20th century years. The year 1917 is considered to be the driest of the 20th century in Texas, mainly because of the lack of precipitation in

Southeast Texas, a region annually receiving a large amount of precipitation relative to the rest of Texas and to New Mexico.

Severe droughts such as those of the 1930s and 1950s have become less common. Groisman et al. (1999) argue that the climate in the United States, particularly in the Southern United States, has shifted to one in which more total precipitation is derived from extreme events. This shift in extreme events also is indicative of a shift in mean precipitation. If mean precipitation is becoming larger, this would imply less frequent and less severe droughts. Gershunov (1998) showed that 50% of a normal station's precipitation falls on less than 20% of the rain days. This shift toward more extreme precipitation events might mean most locations have a higher average precipitation, but this could have implications on drought if the variability of rainfall becomes more extreme.

3. DATA

The COOP observations are recorded on paper forms (1 sheet per month) and are sent to NCDC at the end of each month (Kunkel et al. 2005). These data are subjected to internal consistency checks, compared against climatological limits, checked serially, and evaluated against surrounding stations (NCDC 2006). Special quality control was done on the data pre-1948. The TD-3206 (COOP pre-1948) database for the rescued data is separate from the standard TD-3200 (COOP) cooperative network database (NCDC 2005). The data underwent double-keying, which seeks to minimize the number of keystroke errors by comparing the discrepancies between the two typists. This does not entirely eliminate keystroke errors, nor does it eliminate problems due to illegibility of the form (Kunkel et al. 2005). For all COOP data, Table 3.1 shows the data quality flags representing the possible derivation of daily COOP values found in the TD-3200 and TD-3206 datasets (COOP 2006).

TABLE 3.1. List of possible derivations of COOP values other than the actual value recorded by the observer.

Substituted TOBS for TMAX or TMIN
Time shifted value
Precipitation estimated from snowfall
Transposed digits
Changed units
Adjusted TMAX or TMIN by a multiple of 10 degrees
Changed algebraic sign
Moved decimal point
Rescaling other than F, G, or H
Subjectively derived value
Extracted from an accumulated value
Switched TMAX and/or TMIN
Switched TOBS with TMAX or TMIN
Substitution of “3 nearest station mean”
Switched snow and precipitation data value
Added snowfall to snow depth
Switched snowfall and snow depth
Precipitation not reported; estimated as “O”
Manually edited value
Failed internal consistency check
Failed area consistency check (beginning Oct 1992)

The COOP data are subdivided into climate divisions, regions determined where climate or agriculture can be considered relatively uniform, though some western divisions place a particular emphasis on drainage basis (Guttman and Quayle 1996). The COOP climate division system dates back to the early twentieth century and has evolved to its current format. One hundred and six climatology sections were established in 1912 for the publication of summaries of data through 1910, though the boundaries of these sections were based primarily on mailing issues rather than homogeneous climate considerations. By the 1940s, climatologists adopted the United States Agricultural Bureau of Agricultural Economics crop reporting districts as its new divisions, a system based on the relationship between crop type and climate. In the mid 1950s, state climatologists realigned some of their divisional boundaries to better suit their needs (US Department of Commerce Weather Bureau 1958).

By 1965, the current climate division setup was in place with the western divisions realigning to account for the drainage basins (Guttman and Quayle 1996). Today, there are 344 climate divisions in the contiguous United States, with each state subdivided into as many as 10 climate divisions. The division boundaries generally coincide with county boundaries, except in the western part of the United States (Guttman and Quayle 1996). The COOP data is quality controlled by NCDC though the quality control is not as rigorous as that applied to the USHCN data.

The data of interest in this study are COOP data obtained from NCDC for Texas, New Mexico, and climate divisions that border these states, including stations in Arizona, Arkansas, Colorado, Louisiana, New Mexico, Oklahoma, and Utah; 33 climate divisions in all. The period of interest spans the whole length of the COOP program, going back to 1900. In a few isolated instances, some data exists in the two decades before the program was started, but this older data will not be used here.

The USHCN is a high-quality dataset of monthly averaged maximum, minimum, and mean temperature and total monthly precipitation developed to assist in the detection of regional climate change (Quinlan et al. 1987). The stations were chosen using a number of criteria including length of period of record, percent missing data, number of station moves and other station changes that may affect the data homogeneity, and spatial coverage (Karl et al. 1990). The data for each station in the USHCN are subjected to several steps of quality control, homogeneity testing, and adjustment procedures. The quality control process eliminates outliers while adjustment techniques look at the discontinuities in the different times series. The homogeneity testing looks at time of observation bias, instrumentation changes, and the effects of urban warming.

An important and rather obvious difference between the COOP and USHCN data sets is the number of available stations. Though the two data sets are being treated differently, in reality, the USHCN, as described previously, is a more-rigorously controlled subset of the COOP data. Because of the rigorous quality control already applied to the USHCN data, the stations in this network will serve as the baseline homogeneous station for interpolation of missing COOP data.

However, the geographical coverage of the USHCN stations is poor considering these stations are a subset of the COOP data, with only 1,221 COOP stations of the more than 23,000 available COOP stations chosen. Figure 3.1 shows the geographical location of the USHCN stations in Texas, New Mexico, and surrounding states. Appendix A has detailed information about all 221 USHCN stations in Arizona, Arkansas, Colorado, Louisiana, New Mexico, Oklahoma, Texas, and Utah including station ID, name, latitude, longitude, and elevation.

Our specific months of interest are the 1,224 months from January 1900 through December 2001. The climatic element of interest in these files is precipitation (PRCP), though procedures for the quality control of maximum temperature (TMAX) and minimum temperature (TMIN) will be described. Also of interest in any study on drought is radiation, wind, and relative humidity, but these elements have poor spatial and temporal coverage over the last century.

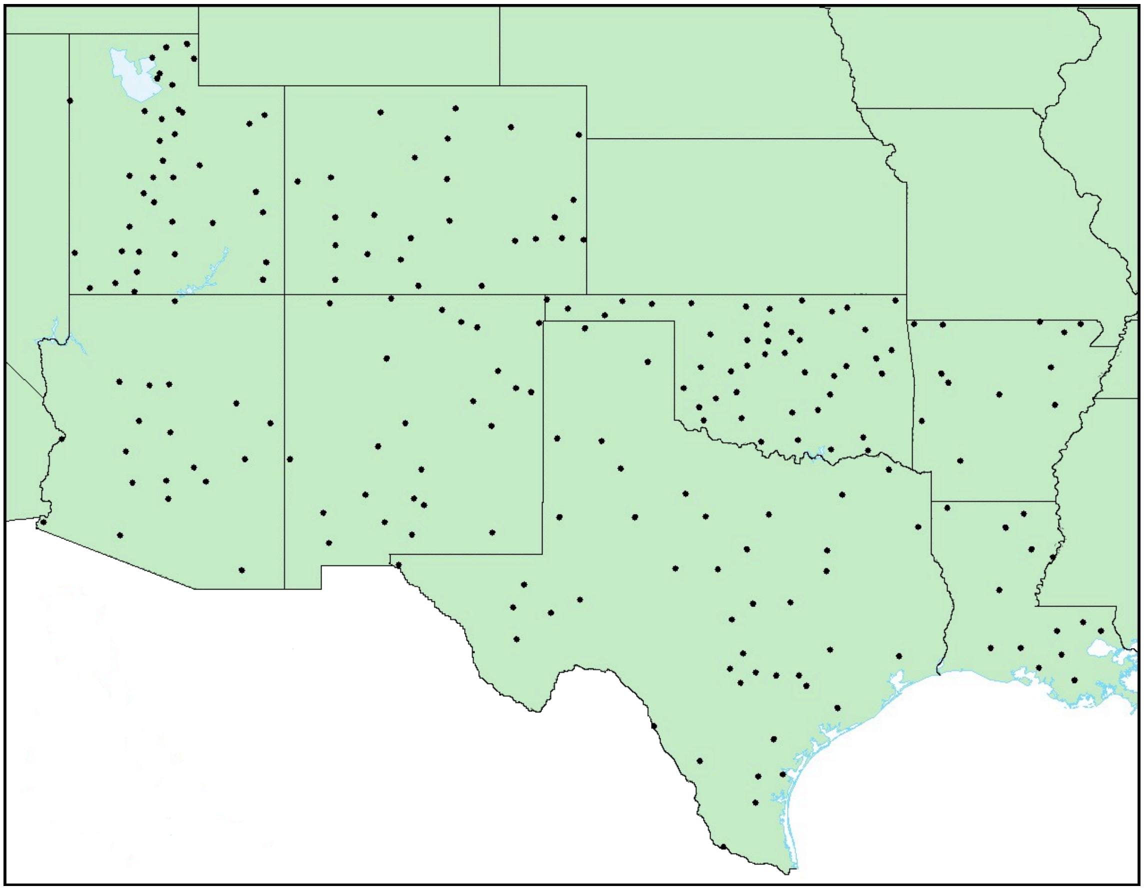


FIG 3.1. The locations of the 221 USHCN stations in Arizona, Arkansas, Colorado, Louisiana, New Mexico, Oklahoma, Texas, and Utah.

The USHCN data were obtained from NCDC via an ftp download and include all the contiguous states. Each state has its own file containing all the data available for that state. The variables available in the USHCN data set are the same as those in the COOP data. The USHCN data set is a quality controlled database that was adjusted substantially to account for a variety of potential contaminants to the data set (Karl et al. 1990). Because the stations used in the USHCN were taken from COOP data, there is some overlap of the data. However, the USHCN station data have a much higher degree of expected accuracy due to the quality control and adjustments.

A simple constraint was put on the COOP data to eliminate stations with incomplete data records. Only stations with at least five monthly values for each month of the year were used. More plainly, a station must have at least five January data, at least five February data, and so on for all twelve months. The interpolation method that will be described in detail is a data driven scheme that works best with a serially complete time series. However, very few COOP stations have serially complete time series, so testing done on the standard errors of values at USHCN show then interpolation process deteriorates significantly when less than five years of data are available.

For each station, this check is done separately between the PRCP, TMAX, and TMIN variables. Table 3.2 shows the number of USHCN stations available by state and Table 3.3 shows the number of COOP stations for each climate division. Each USHCN station contains data for the PRCP, TMAX, and TMIN variables, but several COOP stations have only precipitation data, and very few have only temperature data.

TABLE 3.2. The total number of USHCN stations in each state.

State	**Number of Stations**
Arizona	19
Arkansas	13
Colorado	25
Louisiana	14
New Mexico	24
Oklahoma	44
Texas	44
Utah	38
Total	**221**

TABLE 3.3. The total number of COOP stations in each climate division.

State	**Climate Division**	**Number of Stations**		
		PRCP	**TMAX**	**TMIN**
Arizona	2	85	72	72
Arizona	7	98	66	66
Arkansas	7	32	13	13
Colorado	1	99	66	65
Colorado	2	111	98	97
Colorado	5	23	21	22
Louisiana	1	37	10	10
Louisiana	4	19	9	9
Louisiana	7	36	19	19
New Mexico	1	34	31	33
New Mexico	2	86	54	52
New Mexico	3	58	33	33
New Mexico	4	41	29	30
New Mexico	5	30	21	20
New Mexico	6	39	26	26
New Mexico	7	58	44	44
New Mexico	8	50	30	30
New Mexico	**Total**	**396**	**268**	**268**
Oklahoma	1	29	16	16
Oklahoma	4	24	15	15
Oklahoma	7	26	15	15
Oklahoma	8	33	21	21
Oklahoma	9	30	16	16
Texas	1	94	56	55
Texas	2	94	43	43
Texas	3	215	94	94
Texas	4	135	69	69
Texas	5	91	52	52
Texas	6	143	54	55
Texas	7	102	57	57
Texas	8	66	40	40
Texas	9	50	30	30
Texas	10	13	12	11
Texas	**Total**	**1003**	**507**	**506**
Utah	7	44	40	38

A quick glance at the table suggests that there is approximately the same number of stations meeting the five-year criteria for both maximum and minimum temperature. Every climate division contains more stations with precipitation data than temperature data. Figure 3.2 is a time series graph containing the number of total COOP stations with precipitation, maximum temperature, and minimum temperature data from 1900-2001 for the 33 climate divisions listed in Table 3.3.

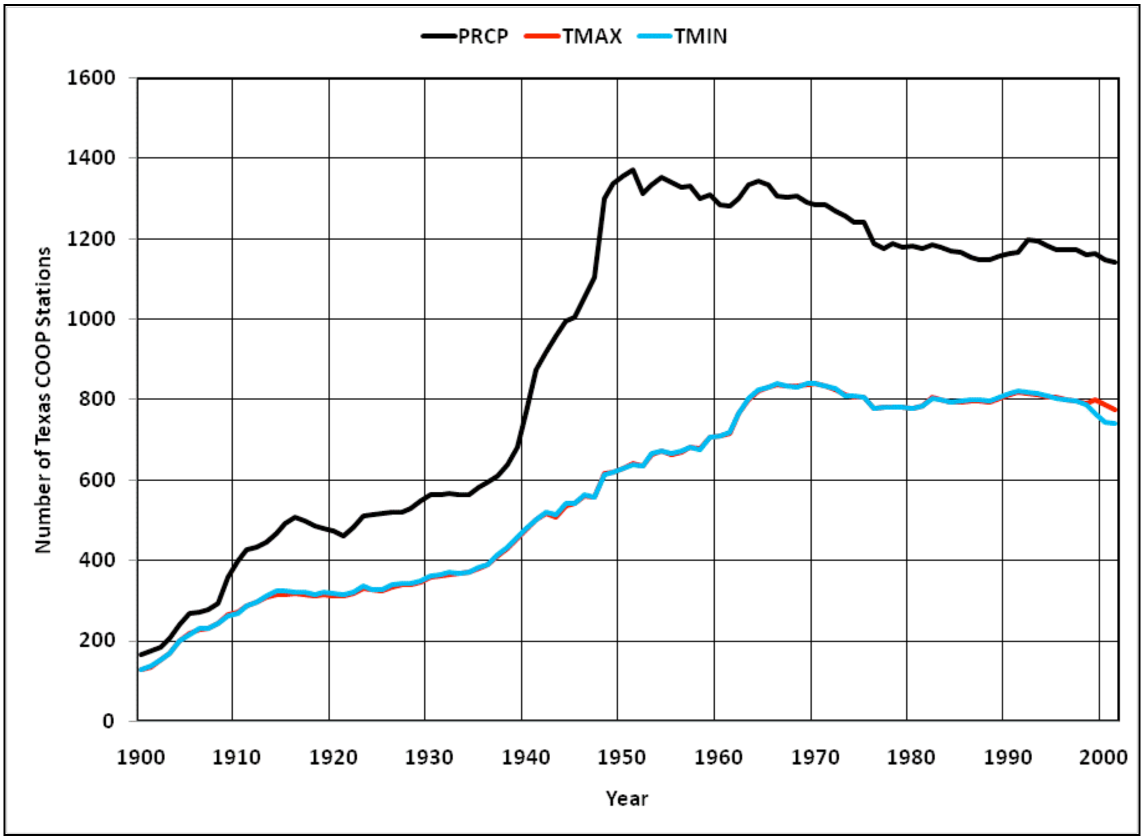


FIG 3.2. The total number of COOP stations from the 33 climate division as a function of time. Specifically shown are precipitation (black), maximum temperature (red), and minimum temperature (blue).

4. QUALITY CONTROL OF COOP DATA

a. Introduction

Throughout this section, quality control of individual stations will be described. A target station throughout this study will be defined as a station's monthly time series to which a process being described is applied. Neighboring stations refer to the set of available stations other than a target station. The rest of this paper will refer to target and neighboring stations in this manner. When a process is described for a target station, this process is repeated for all the COOP stations.

The raw COOP data received from NCDC were in the form of daily precipitation totals (PRCP) and daily maximum (TMAX) and minimum (TMIN) temperatures. The daily values for each COOP station were sorted to eliminate extreme values indicative of missing data or unrealistic values. An example of an unrealistic value would be a negative value for precipitation. Monthly values (averages) for COOP stations were only calculated when all possible days of a month were present and did not contain unrealistic values or missing data flags.

There are a number of potential sources of errors or quality issues in the COOP dataset that are separate from data flag issues. These errors generally fall into three categories: observer error, station discontinuity, and digitization errors (Kunkel et al 2005). Observer errors include errors in reading the instruments or in writing the observations on the form, and problems with the equipment (Kunkel et al 2005). Station discontinuity errors are caused by gaps in the record of a station, station moves, and changes in the landscape around the observation equipment. Something as simple as an observer taking a vacation or leave each year at a particular time can leave undesirable gaps in a station's data set. Digitization errors may include misidentifying stations and keystroke errors, which may stem from difficulty in reading the original form submitted by an observer (Kunkel et al 2005).

After the elimination of the missing data flags and unrealistic values from the COOP daily values, further quality control checks were performed on the monthly precipitation totals and monthly temperature averages derived from the daily COOP data. This quality control check compared stations within the same climate division and searched for outliers for each available month of data based on the other available stations within its climate division. This process was performed separately for precipitation, maximum temperature, and minimum temperature data. For instance, it is possible for a station with a monthly average flagged as an outlier for minimum temperature to not have its maximum temperature flagged as an outlier.

b. Eliminating Missing and Unrealistic Values

The standard flag in daily COOP data for a missing daily value is -99999. For fixed length records only, when a data value is missing, the sign of the data value is set to "-", the data value is set to "99999", flag position 1 is set to "M" and flag position 2 is blank. Flag position 2 refers to the lettering scheme which labels the derivation of an official daily COOP station value other than the daily value recorded by an observer. For variable-length records, the minus sign is omitted for any such values (NCDC, 2006). This quality control work on the COOP data failed to account for another extreme value, "-99", which often was present in the daily COOP data sets.

The flag “-99” may lower monthly precipitation values to what appears to be a reasonable value, since the raw data file contains precipitation in hundredths of inches. Also, because temperature is an average of approximately thirty daily values, one or two of these flags may lower the temperature to a reasonable, but inaccurate, monthly average. In order to eliminate these flags from the data, the maximum and minimum values for each month of available data were listed. If the minimum value for a particular month was “-99” or “-99999”, that month of data was eliminated. Likewise, if a maximum value was “99999” or some another unreasonable number, that particular month was eliminated.

Another issue in the daily COOP data files is the existence of missing data flags for months that have less than 31 days. The 31st data value is always represented by a missing data flag, typically “-99999”. Therefore, in months that have less than 31 days, the 31st missing data flag was ignored in the extreme value check just mentioned. Depending on whether or not a particular year was a leap year, the 29th, 30th, and 31st missing data flags were ignored in February on non-leap years and the 30th and 31st on leap years. An algorithm was developed to ensure that the missing data flags representing these “days” were eliminated.

M*icrosoft Excel* sorting functions were used to eliminate the months containing missing data flags. However, this check did not eliminate potential outliers for data values unless they represented missing data flags or unrealistic values. The only unrealistic values eliminated in this process were negative precipitation values and high (low) temperatures more than 50 degrees more (less) than the closest extreme value.

c. Creating Residual Values Based on Other Stations

Because of the relative homogeneity within climate divisions, quality control for errors described by Kunkel et al. (2005) was performed by intra-comparison of climate division data. For each available monthly value for a target station, a comparison value was created using the other stations with the target station's climate division with available, valid data for that particular month and year. Thus a mean value of the available monthly values in a target station's climate division can be calculated. This mean value represents a monthly value or average "typical" for that given month in a target station's climate division.

The given COOP target station value for a month and year are subtracted from this mean value for the same month to get a residual value. Across the target station's time series, residual values were created for all available months. These residual values were only created when at least four other stations had valid data in a given month in a target station's climate division.

Figure 4.1 shows the values for stations in a climate division used to calculate this mean value for precipitation. The graph shows monthly precipitation values in climate division 3 for both August and September of 1978. These values correspond to the quality control check of COOP station 410120 for both months. August 1978 was a particular month in which station 410120 would eventually be flagged in the quality control process. September 1978 was a month for which data from 410120 was accepted.

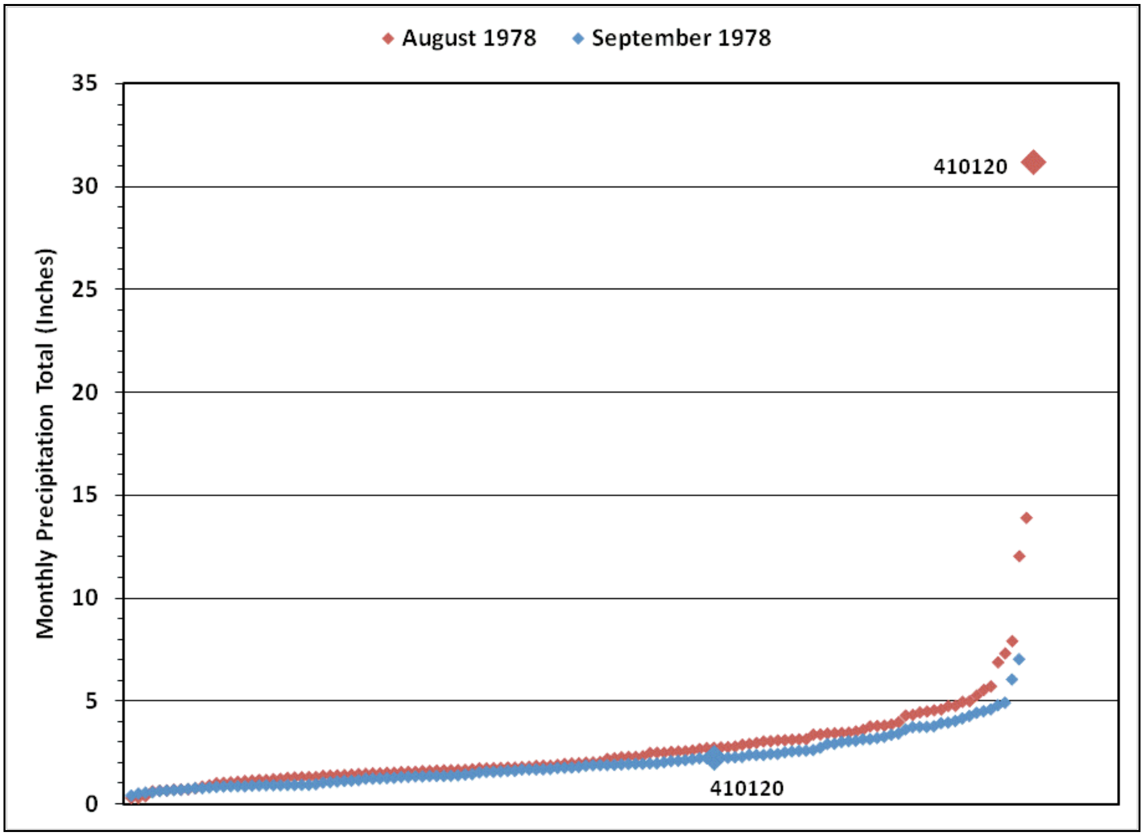


FIG 4.1. Monthly precipitation values in Climate Division 3 for August 1978 (red) and September 1978 (blue), with the large diamonds representing COOP station 410120.

The monthly total of precipitation recorded for COOP station 410120 for August 1978 was 31.19", far more than any other station with a monthly value for that month in Texas climate division 3. The mean value according to the precipitation values other than COOP station 410120 was a comparatively small 2.62". Figure 4.1 also shows the monthly precipitation values for September 1978 in climate division 3 for COOP stations other than 410120. The mean monthly total for these 126 stations is 2.12", which compares very favorably to the recorded total of 2.25" for station 410120 in this month.

Figure 4.2 shows the values for stations in a climate division used to calculate this mean value for temperature. More specifically, these are the monthly average maximum temperature values in climate division 8, which contains COOP station 417186 in Port

O'Connor, TX. The values of the other stations in this climate division are those used to test the validity of the value at COOP station 417186. The series contains data from April 1996, a particular month in which COOP station 417186 was flagged in the quality control process. The second series shows data for the following April, a month that passed the quality control check according to the other data for this month in climate division 8.

The April 1996 monthly average maximum temperature at COOP station 417186 of 69.27°F was far less than any other station in climate division 8. The mean value according to the average values other than COOP station 417186 was much larger, specifically 79.54°F. The average maximum temperature of 72.37°F at station 417186 the following April was also relatively low compared to the rest of the region (74.47°F). However, the distribution of temperature shows the April 1997 maximum temperature average for this station to be a reasonable value.

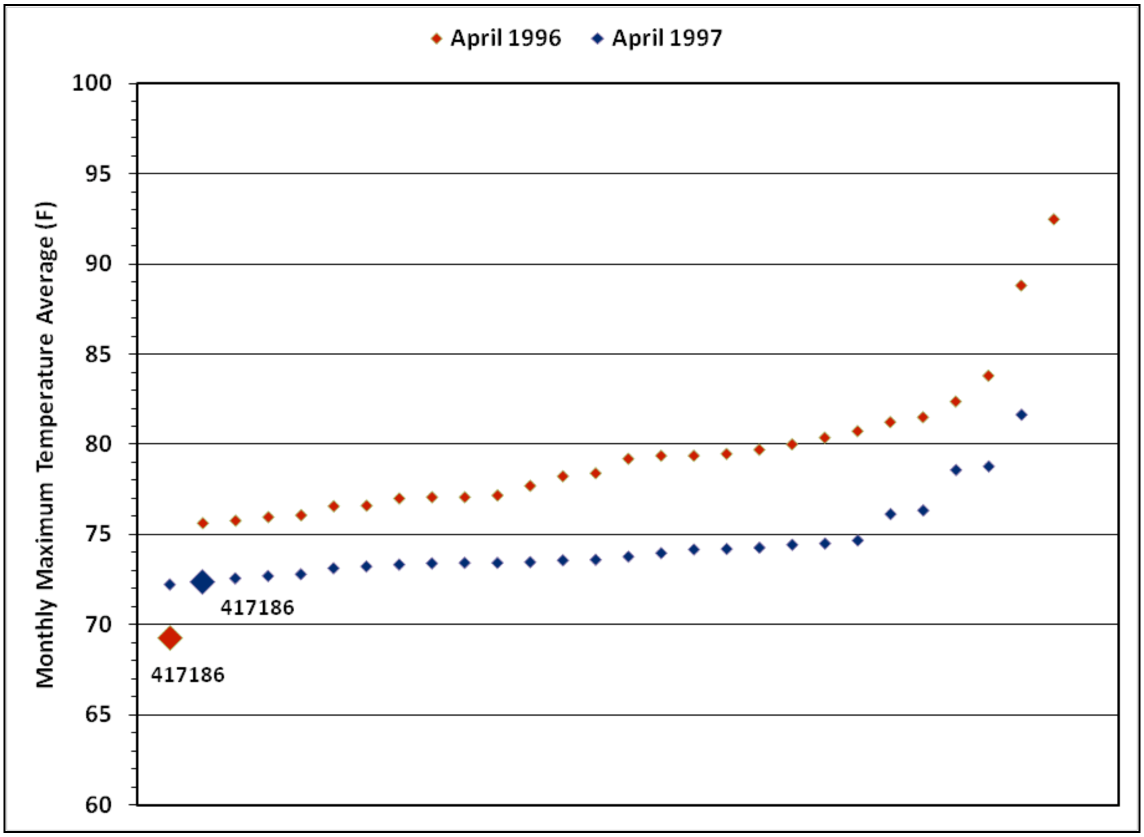


FIG 4.2. Monthly average maximum temperature in Climate Division 8 for April 1996 (red) and April 1997 (blue), with the large diamonds representing COOP station 417186.

d. Quality Control Check of Residual Values in Time Series

These resulting residual values for a target station's time series were then ranked from lowest to highest in order to find the 25th and 75th percentile values. The ranking procedure following is based on González-Rouco (2000), which used a rank percentile system to find outliers. Then, the difference between the 25th and the 75th percentiles is taken to be the inter-quartile range (IQR). This rank analysis is done for all the available monthly data and includes all months of the year in the same analysis.

Figure 4.3 is a time series graph of the residuals for COOP station 410120. For each month, the residual is calculated as the difference between the recorded value at station 410120 and the mean value derived from the rest of the available COOP stations in Texas climate division 3. Likewise, the residuals for COOP station 417186 are shown in Figure 4.4, calculated from the other stations in climate division 8.

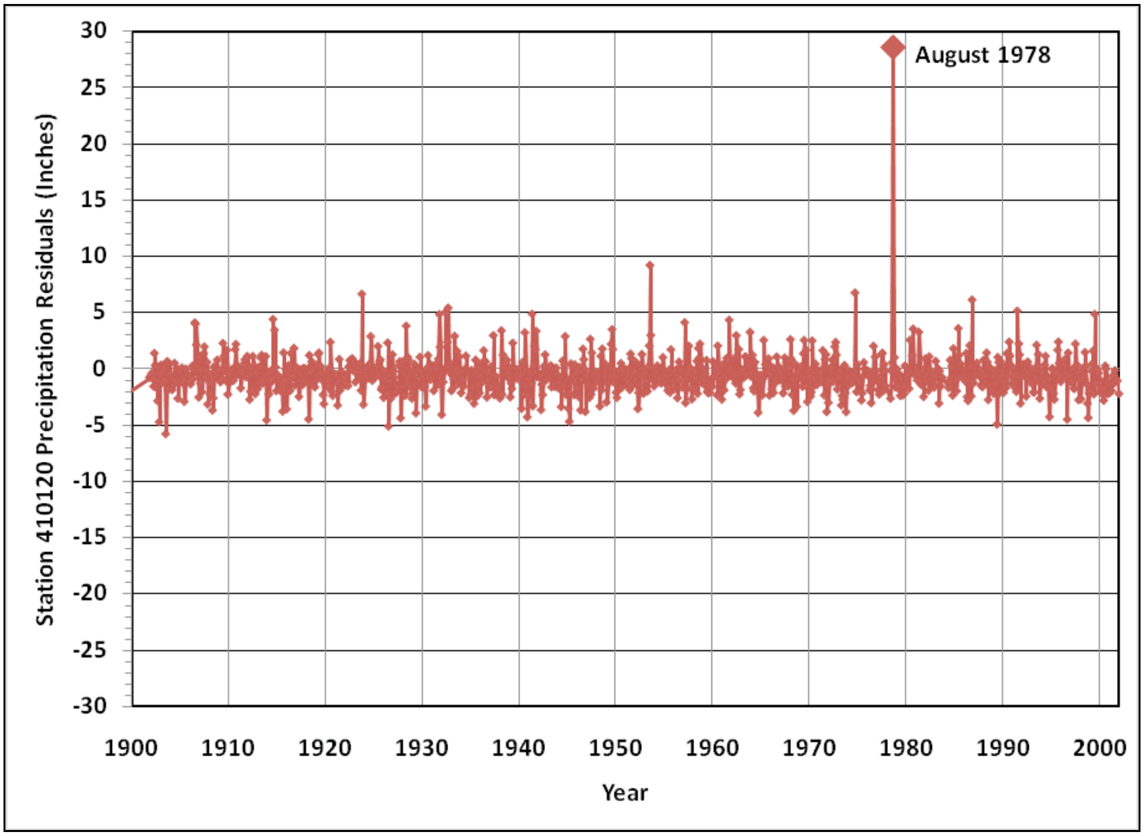


FIG 4.3. Time series of monthly precipitation residuals for COOP station 410120 from 1900 through 2001. The large diamond represents the residual from August 1978.

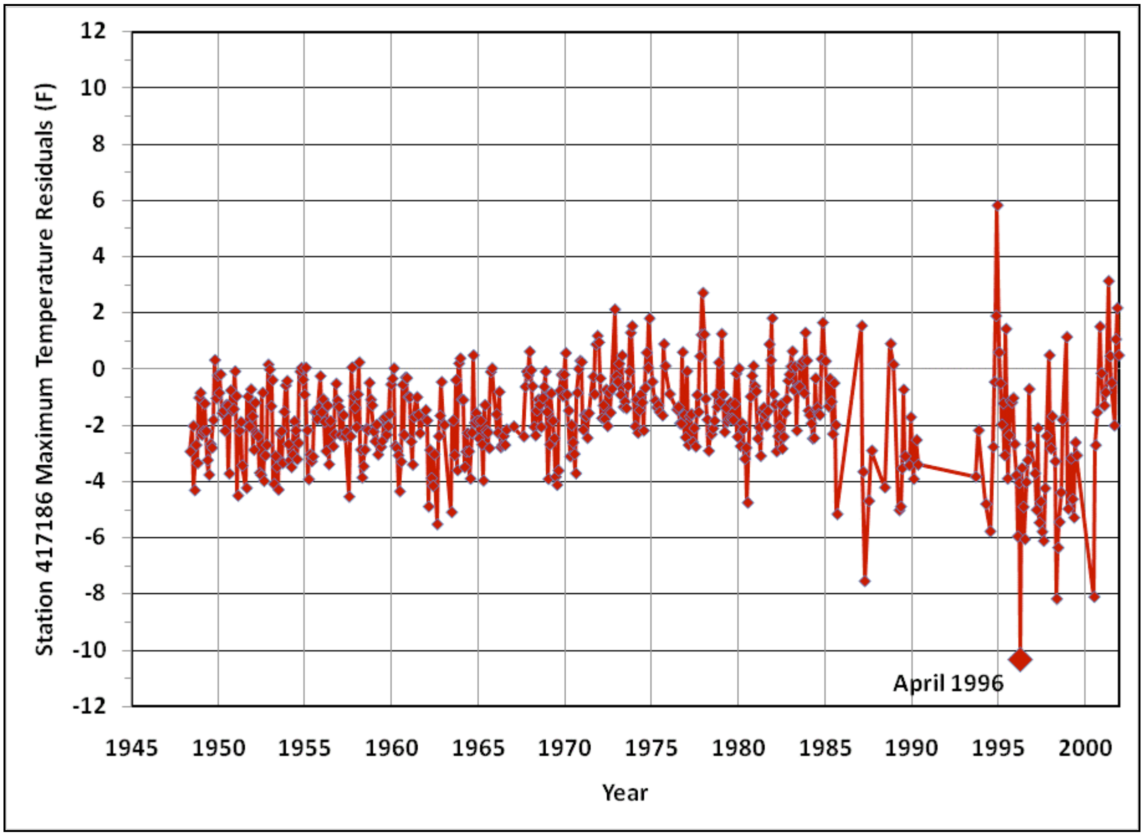


FIG 4.4. Time series of maximum temperature monthly average residuals for COOP station 417186 from 1945 through 2001. The large diamond represents the residual from April 1996.

The residual for August 1978 monthly precipitation is prominent on Figure 4.3, and although there are a few other spikes, the magnitude of this residual by far the largest in this time series. Another characteristic of this time series of residuals is the lack of spikes with significant magnitude with negative values. For a large negative spike to occur, the mean precipitation would have to be extremely high while the recorded monthly value at a COOP station is near zero. This scenario is highly unlikely because it would indicate very isolated areas with little to no precipitation surrounded by a large area with unusually high precipitation totals. This implies the quality control check will eliminate months with large positive anomalies.

Unlike the residual time series plot for COOP station 410120 PRCP, the residual time series of 417186 TMAX has two distinct spikes, both for the month in question and for December 1994. Also, station 417186 has a period of record about half the length of station 410120 and has fewer stations within its climate division for comparison. Further analysis of the positive spikes shows the vast majority to occur in the winter months, an expected occurrence considering COOP station 417186 is a coastal station that is directly next to the Gulf of Mexico.

The winter months in climate division 8 are examples of a small area having a different climatology than the remainder of the climate division. The quality control procedure works under the assumption of a relatively homogeneous climate within a climate division, so this coastal influence is not accounted for. Outside the winter months, the recorded average maximum temperature value is generally less than the climate division 8 mean. The largest magnitude for a minimum spike in the COOP station 417186 residual plot occurred in the month of April 1996.

Both the precipitation and temperature residual values used in this test are assumed to be a normal distribution centered on zero. This assumption holds up well for temperature values, but not as well for precipitation values. We showed earlier that this test can incorrectly eliminate months containing unusually large positive anomalies of precipitation.

After the IQR for the residuals in the time series is calculated, we find the values in the time series identified as outliers by González-Rouco (2000), namely the 25th percentile – 3 IQR and 75th percentile + 3 IQR. This process identifies both uncharacteristically high and low error values in a given precipitation or temperature time series. Figure 4.5 is a scatter plot of the times series residual value with the IQR, the 25th percentile – 3 IQR, and the 75th percentile + 3 IQR lines denoted for COOP station 410120.

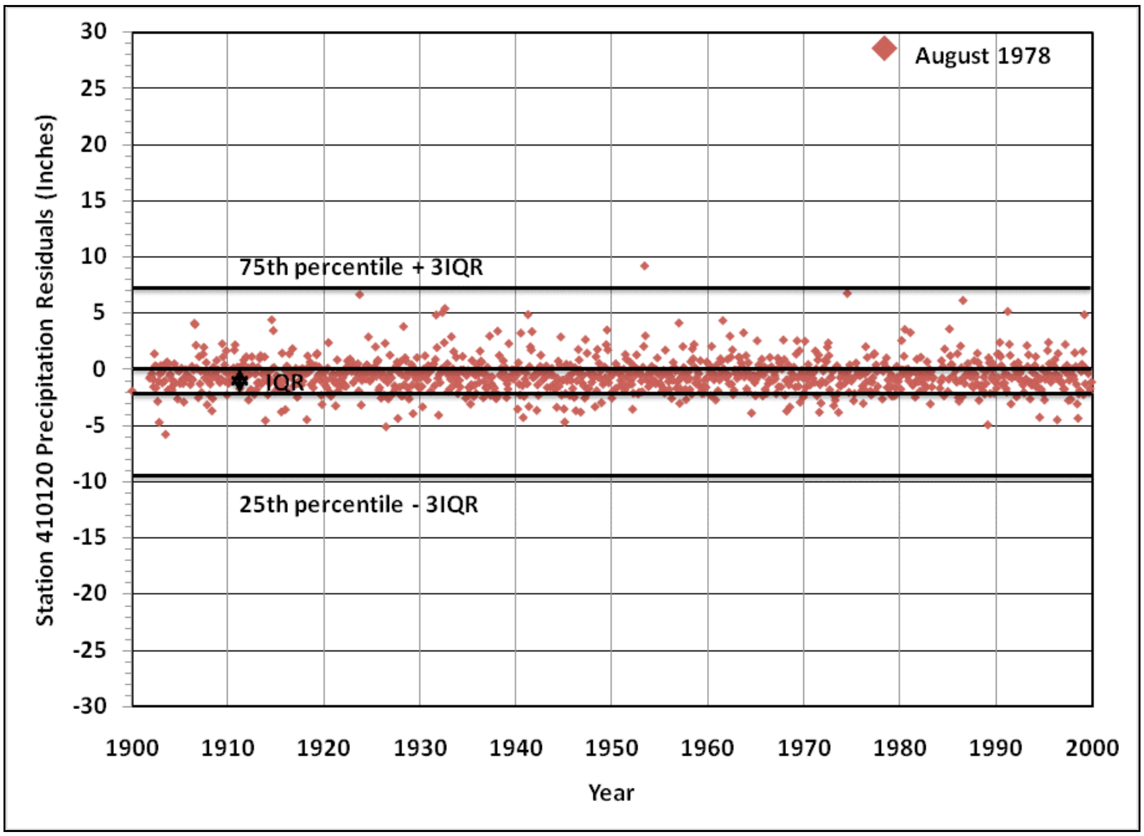


FIG 4.5. Scatter plot of monthly precipitation total residual values for COOP station 410120 from 1900 through 2001. Included are lines for the 25th percentile and 75th percentile and the values representing 3IQR + 75th percentile and 3IQR - 25th percentile are also included.

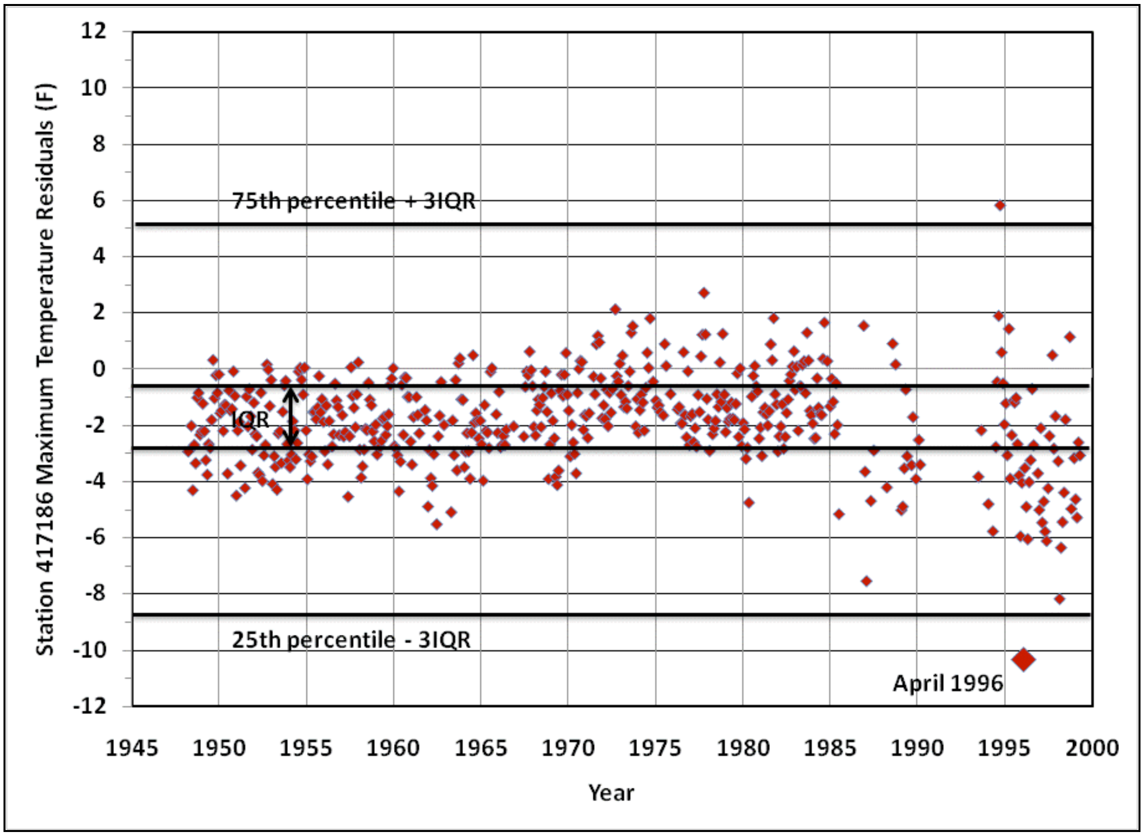


FIG 4.6. Scatter plot of maximum temperature monthly average residual values for COOP station 417186 from 1945 through 2001. Included are lines for the 25th percentile and 75th percentile, and the values representing 3IQR + 75th percentile and 3IQR - 25th percentile are also included.

Figure 4.5, the graph of 410120 PRCP residuals, indicates that all of the outliers based on the time series of the residuals lie above the 75th percentile + 3 IQR line. The 25th percentile for this time series has a residual of -1.40” while the the 75th percentile has only a positive magnitude of 0.09”. This would suggest that COOP station 410120 is usually drier than the mean value of the surrounding stations for a particular month.

Figure 4.6, the scatter plot of 417186 TMAX residuals, shows that the IQR range is about 2.00°F, with mostly negative residuals indicating a cooler than normal climate. However, there are two outliers according to the IQR test based on González-Rouco

(2000). These two outliers are the spikes indicated earlier in the time series graph of residuals, the negative outlier representative of April 1996.

The step of the quality control process looking at monthly residuals ensures that climatologically warm, cool, wet, or dry stations do not have data eliminated because of a systematic difference from the climate division average. For instance, the 50th percentile error value for a particular station may be +4.25°F. The residual time series quality control is done to ensure the same general climate trends are taking place in the target station as in the stations used to create the error values.

e. Quality Control Check Against Other Stations Within Climate Division

If the data for a particular month at a target station is deemed an outlier based on its own time series residuals, a check is performed against the other available stations in its climate division. For each month and year the data is flagged, the other available station data in the climate division of a target station are ranked from highest to lowest in order to calculate an interquartile range (IQR_2 to avoid confusion). This interquartile range (IQR_2) is different than the previous interquartile range (IQR) calculation and is based only on climate division data for that given month and year. For a month flagged in the previous quality control check, there is now a ranking of station data in a target station's climate division, less the target station. Figure 4.7 is a scatter plot shows the monthly precipitation totals from climate division 3 less COOP station 410120 for August 1978, along with IQR_2. Figure 4.8 is a scatter plot of the average monthly maximum temperatures from climate division 8 without COOP station 417186.

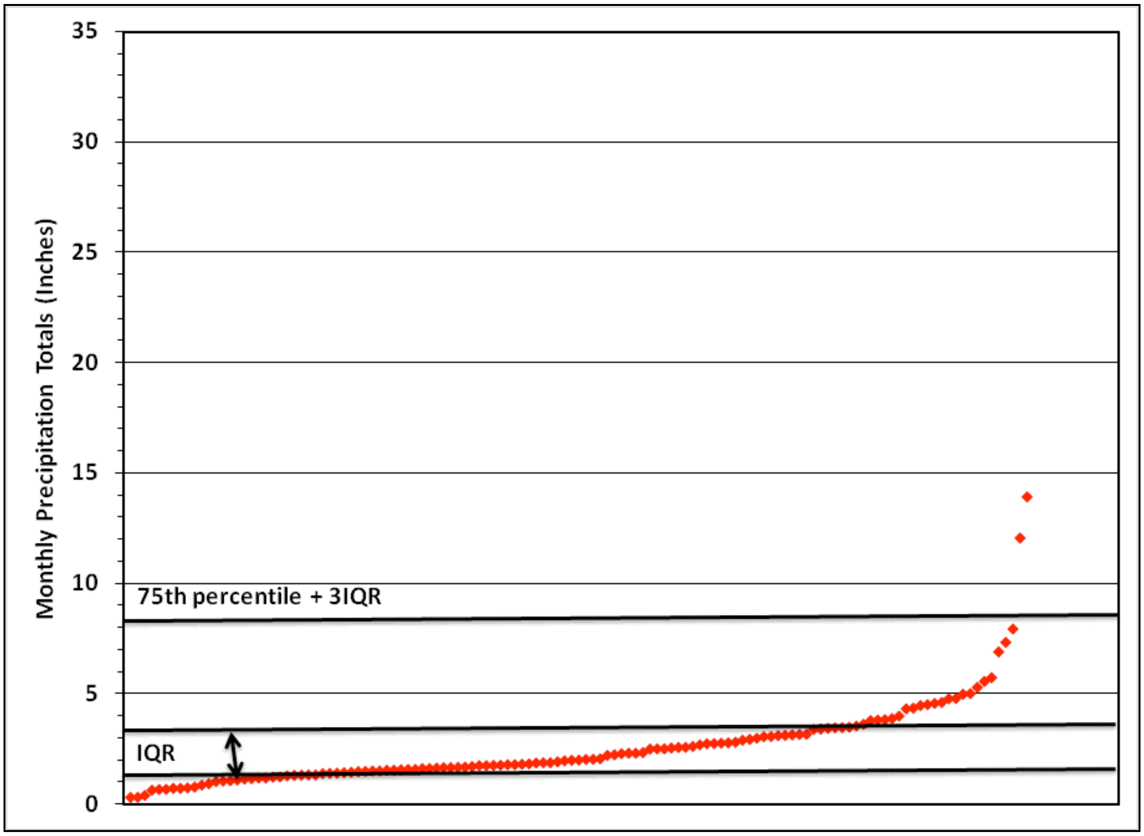


FIG 4.7. Scatter plot of August 1978 monthly precipitation totals for COOP stations in Climate Division 3 without station 410120.

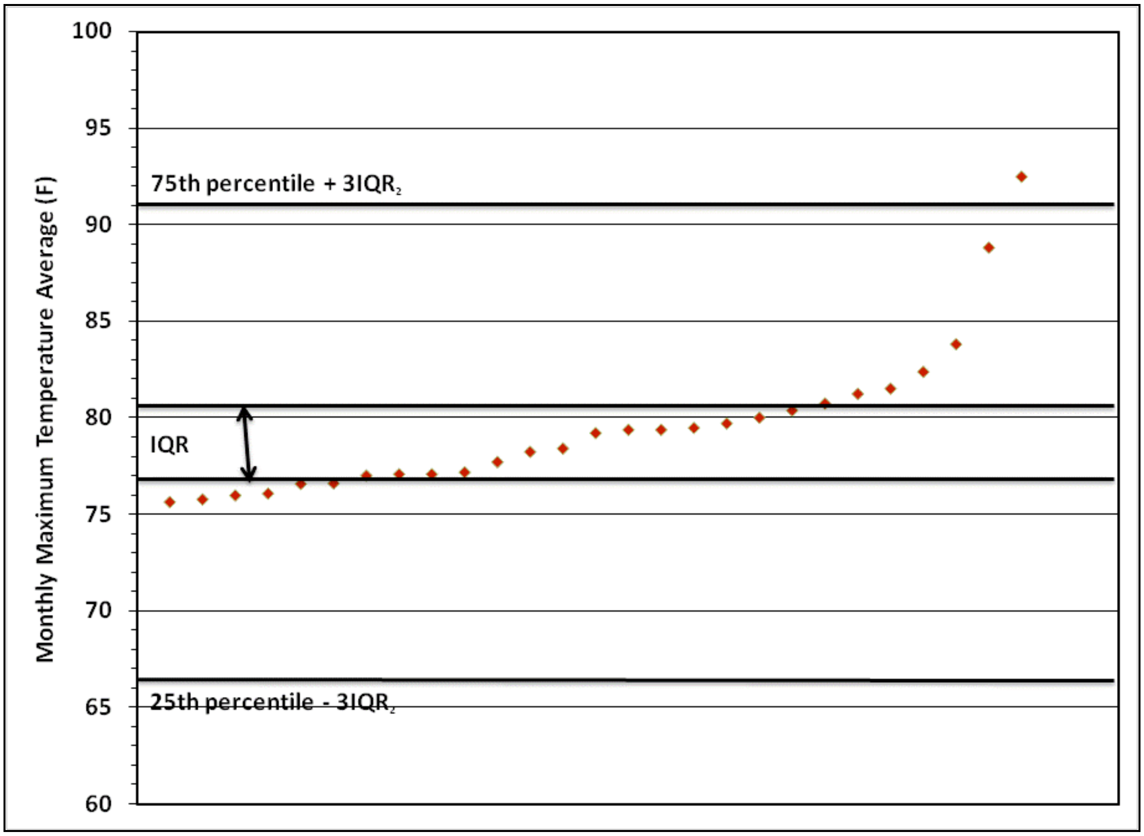


FIG 4.8. Scatter plot of April 1996 maximum temperature averages for COOP stations in Climate Division 8 without station 417186.

Data flagged as suspicious based on the residual time series check are further analyzed to see if their elimination as unrealistic values is warranted based on two further criteria. The first criterion says if its actual value (not error value) is less than the 25th percentile of the values for that month - $3IQR_2$ or the 75th percentile of the values for that month + $3IQR_2$, in this case 11.70”. The values for a climate division for a specific month are all the available stations except for the target station.

The example of COOP station 410120 for August 1978 shows that two other stations within the region fell above the line of 75th percentile of the values for that month + $3IQR_2$. The precipitation totals are arranged from smallest to largest and independent of COOP station ID number. The monthly total of 31.19” for COOP station

410120 is much larger than the 75th percentile of the values for that month + $3IQR_2$. For August 1978, Figure 4.7 shows two COOP stations reporting precipitation above this threshold. The two stations other than 410120 were 410478 (13.93") and 419014 (12.07"). Therefore, an additional quality control check was added to ensure a value flagged as suspicious was unique compared to other stations in its climate division.

This additional criterion differs for precipitation and temperature. For precipitation, if the monthly value for the target station is more than $3IQR_2$ added to the 75th percentile, the second criterion is met if the target value exceeds twice the value of every neighboring station in its climate division. Conversely, if the monthly value for the target station is less than $3IQR_2$ subtracted from the 25th percentile, the second criterion is met if the target value is less than half the value of every neighboring station in its climate division. The second criterion for temperature is met if the monthly value for the target station is either five degrees more or five degrees less than every other station available in the division. Again, for temperature, the testing of this second criterion is dependent on whether this extreme value is significantly more or less than the other values within the target station's climate division.

The COOP station 410120 monthly precipitation total of 31.19" in August 1978 is flagged by both the time series test and the test against other station data for that month. The last criterion to be met in order to be completely flagged by the overall quality control test is assuring this data point is at least twice the magnitude of every other monthly value for August 1978 in climate division 3. The bar graph in Figure 4.9 shows the ratio of the monthly total for COOP station 410120 (31.19") to every other precipitation total in its division. According to the quality control checks, the August

1978 monthly data precipitation total recorded as 31.19” is deemed an outlier and eliminated by the quality control checks.

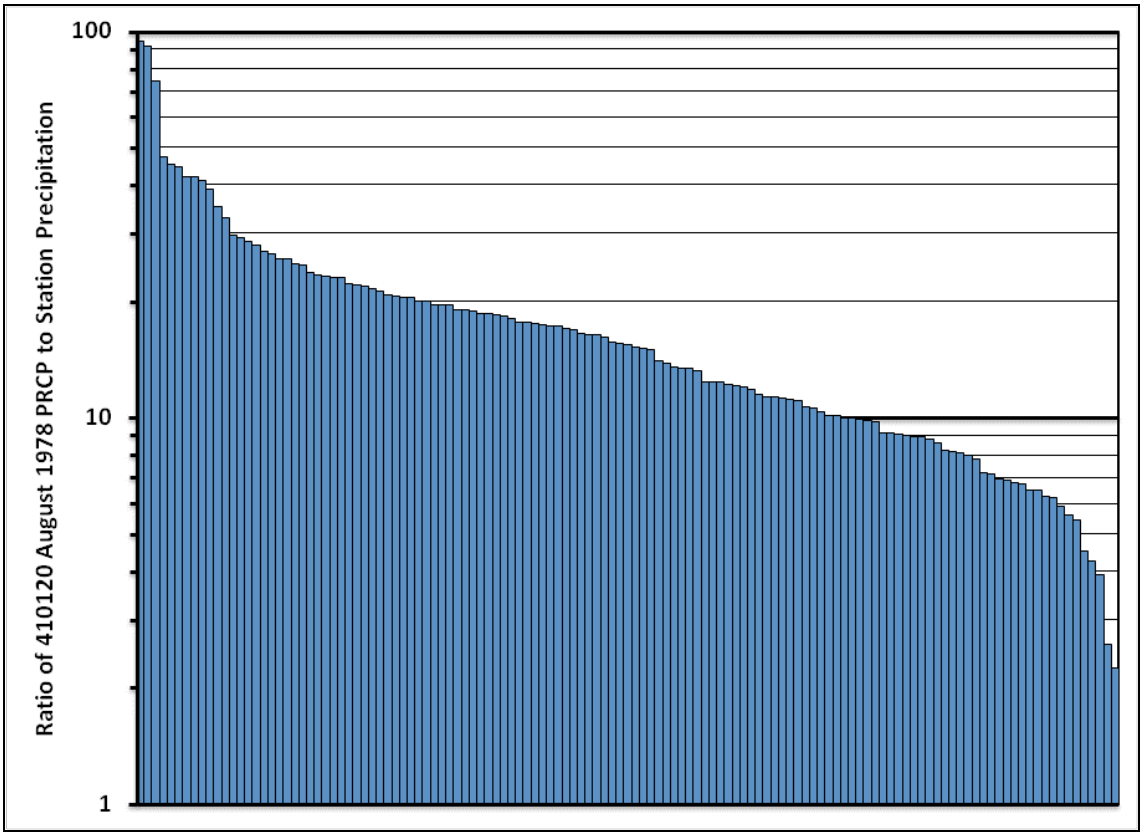


FIG 4.9. Bar graph showing the ratio of the August 1978 monthly precipitation total at COOP Station 410120 to all the other recorded totals in Climate Division 3 (Logarithmic scale).

The final outlier criterion for temperature is that the recorded value is at least five degrees more or five degrees less than every other station for that particular month. The bar graph below shows the difference between the COOP station 417186 monthly maximum temperature (69.27°F) and each of the COOP stations recording monthly maximum temperature values in climate division 8 for April 1996 (Fig. 4.10). The graph shows that the recorded value of 69.27°F is at least five degrees less than every other value recorded in climate division 8 for April 1996.

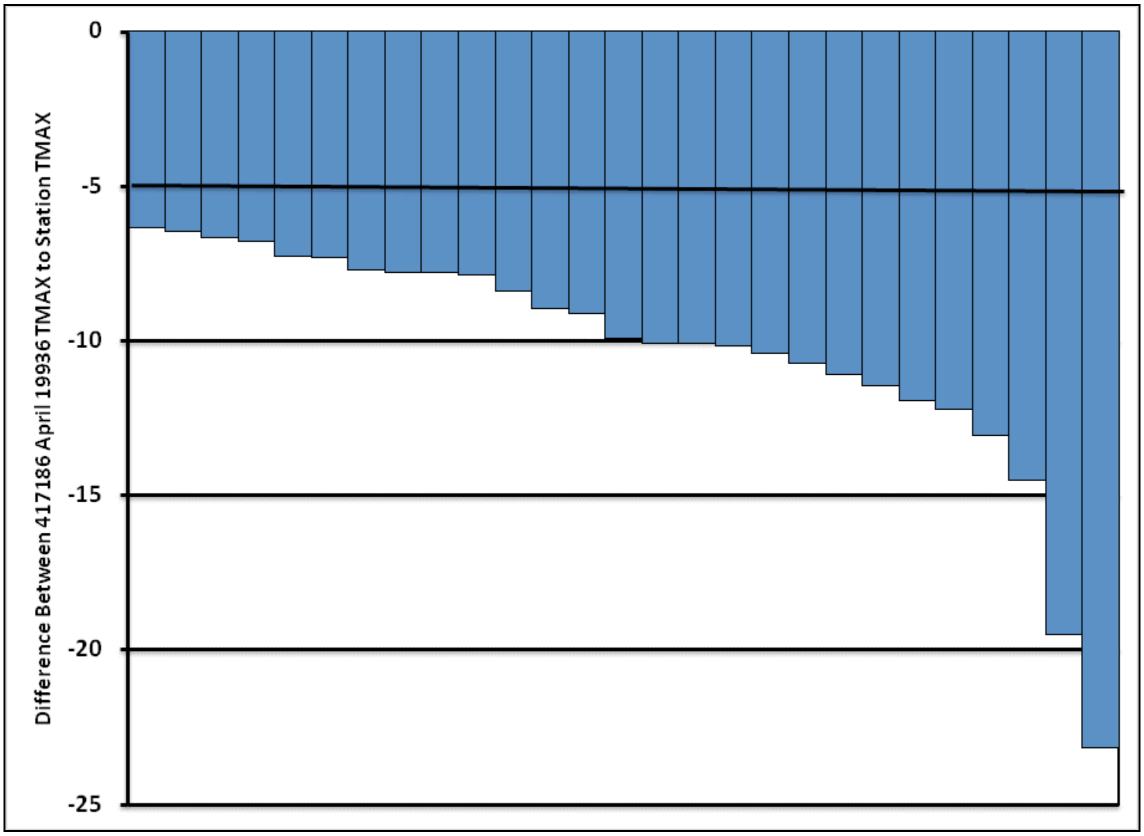


FIG 4.10. Bar graph showing the difference of the April 1996 maximum temperature average at COOP station 417186 to all the other recorded averages in Climate Division 8, including a line for the difference of -5°F.

Monthly precipitation totals and monthly temperature averages are deemed as outliers and flagged if they meet the requirements of the two aforementioned quality control checks. However, an additional quality control check was added to two or more eliminate suspicious data that occurred in a short time period. More specifically, if two or more suspicious data eliminated by the quality control procedures occurred within a five year period, the entire period was thus eliminated. The earliest month of data removed was one year prior to the oldest flagged month of data. Likewise, the last month of data in this period to be removed was a year after the most recent flagged month of data within this period.

The process of eliminating data was modified slightly for data flagged at the beginning of a particular station's climate record. In the case that two or more data were flagged within five years of the start of a station's climate record, all data were removed from the beginning of the climate record through one year after the most recent data flagged in this period.

COOP station 296435 from New Mexico climate division 8 represents a station that had more than one data point eliminated by the quality control check within a five year period. Figure 4.11 is a time series plot of its monthly residuals, defined as the difference between the 296435 value for a given month and the climate division mean for the same month. This particular station is in New Mexico Climate Division 8, the Southern Desert, so the climate division mean for a particular month's precipitation total is often near zero. Summer rains fall almost entirely during brief, but frequently intense thunderstorms in this region (Sheppard et al. 1999).

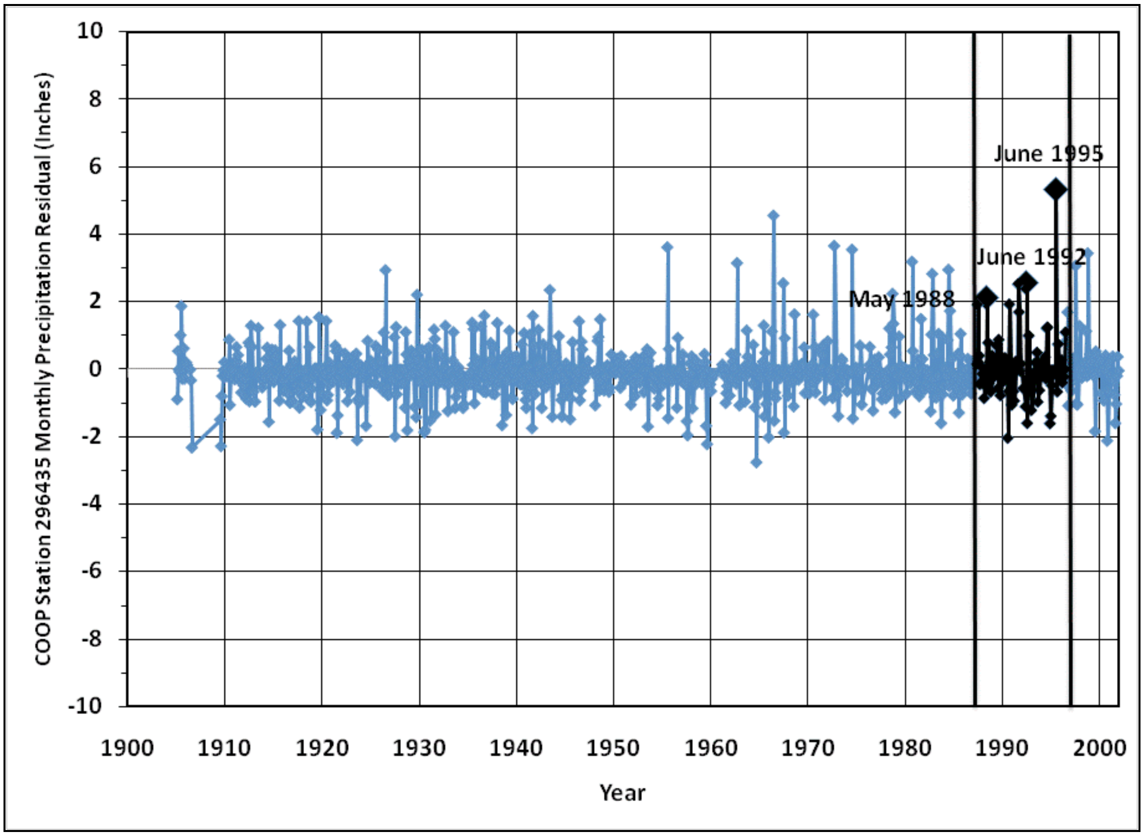


FIG 4.11. Time series of monthly precipitation total residuals for COOP station 296435 from 1900 through 2001. The large diamonds represent months flagged and used to eliminate all data between May 1987 (right vertical line) and June 1996 (left vertical line).

COOP data that passed through the missing and unrealistic data checks and the quality control checks was placed in a dataset called "Quality Control Data." Another dataset was created called "Raw Data," which contains data that only passed the missing and unrealistic data checks. The quality-controlled data will be compared to the second dataset to determine the usefulness of the quality control procedures. In reality, the dataset containing only data that passed the temporal and spatial quality control checks is probably missing some data that was indicative of natural processes, but was seen as an outlier by our quality control processes.

This likely was the case in August 1978 at COOP station 410120 in Albany, TX. Figure 4.7 shows two other stations in climate division 3 to have monthly precipitation totals above the $3IQR_2 + 75^{th}$ percentile threshold for August 1978. 410478 (13.93") is Baird, TX, roughly 25 miles south of Albany and 419014 (12.07") is the COOP ID for Throckmorton, TX, which is roughly 35 miles north of Albany. Further investigation shows this extremely heavy rainfall to be associated with the remnants of Tropical Storm Amelia, a storm that dropped 48" of rainfall in 52 hours at Medina, TX (Dickson, 1978). The total for Medina, located roughly 50 miles northwest of San Antonio in the Texas Hill Country, is the wettest known storm total rainfall amount for both the state of Texas as well as the continental United States (Dickson 1978). Unfortunately, there were nine deaths with the associated flash flooding caused by the heavy rainfall in Albany (Dickson, 1978).

Table 4.1 shows the differences between the "Quality Control Data" and the "Raw Data." The first group of columns represents the number of monthly data elements grouped by climate division available before the quality control checks were applied. These columns can also be interpreted as the number of monthly data elements available in the "Raw Data" dataset for each climate division. The second group of columns represents the number of data elements flagged as missing or unrealistic and the data eliminated by the quality control checks.

TABLE 4.1. Number of data elements checked and flagged for each climate division used in the study.

Climate Region	PRCP		TMAX		TMIN	
	Checked	Flagged	Checked	Flagged	Checked	Flagged
Arizona 2	39855	56	31722	83	31162	57
Arizona 7	47174	339	28025	27	27659	217
Arkansas 7	17960	0	6621	1	6568	0
Colorado 1	44597	233	28007	41	27536	16
Colorado 2	60898	180	47761	65	47217	29
Colorado 5	9979	114	8904	127	8846	43
Louisiana 1	13903	1	5163	0	5180	0
Louisiana 4	7288	0	4258	0	4169	0
Louisiana 7	16333	30	8663	0	8542	0
New Mexico 1	16370	25	13458	130	13309	6
New Mexico 2	44555	54	28019	73	27369	3
New Mexico 3	31461	134	18191	179	18131	81
New Mexico 4	17866	167	9980	160	9990	0
New Mexico 5	13740	410	10966	2	10666	53
New Mexico 6	18391	46	11822	2	11575	58
New Mexico 7	28090	185	18886	7	18581	70
New Mexico 8	26854	253	16737	175	16486	170
Oklahoma 1	13374	56	9434	1	9274	3
Oklahoma 4	16338	6	10606	17	10254	28
Oklahoma 7	18042	5	11586	80	11295	2
Oklahoma 8	18511	113	12423	12	12180	3
Oklahoma 9	15407	1	7062	69	6918	45
Texas 1	55275	52	30163	2	30041	54
Texas 2	47737	6	23263	46	23244	25
Texas 3	108195	6	51556	188	51503	79
Texas 4	64244	92	30592	26	30455	53
Texas 5	38920	456	21168	20	20845	35
Texas 6	56279	133	25146	144	25140	105
Texas 7	54604	192	29231	16	28981	153
Texas 8	38255	62	21035	103	20857	28
Texas 9	27985	124	15151	5	15061	11
Texas 10	9266	137	7257	1	7144	0
Utah 7	18539	75	16924	97	16552	29

5. LITERATURE REVIEW ON INTERPOLATION METHODS

a. Homogenized Datasets

Over the past several years, the homogenization and recreation of climate data has become desirable and led to several hypotheses on how to best solve the problem. Of most interest are approaches that focus on indirect means using data adjustments because of the poor spatial coverage of truly homogeneous data sets. Station data are used in most homogeneity testing techniques but primarily in conjunction with metadata or comparisons with neighboring stations. Using only data from an individual station is problematic because the change (or lack of change) one detects may not be caused (or masked) by real changes in climate (Peterson et al. 1998). However, using data adjustment techniques calls into question the validity of the techniques used (Easterling et al. 1995). Validation of the techniques in most previous studies was done by comparing an interpolated value using strictly surrounding stations to a value that was recorded at a target station. For instance, Keim (2003) used USHCN data to analyze trends in NCDC COOP data.

Scientists that have set out to create homogeneous data sets generally infer missing precipitation and temperature values using several approaches and apply error analyses to objectively determine the best method. Typically, the choice is dependent on several factors: the meteorological variable under consideration; the geographical area; the spatial distribution of surrounding observations; and the month/season for which the target station is to be estimated (Eischeid et al. 1995).

b. Indirect Methodologies for Inferring Data

The following methods are indirect ways of determining values for monthly precipitation totals or average temperature in the case of missing data, unrealistic data, or data that has been removed because of quality control. These methods take advantage of data available at neighboring stations to fill the missing data at target stations.

Perhaps the most basic methodology for inferring station data is an arithmetic mean of two or more surrounding stations for a particular data value. Alexandersson (1986) used three different techniques to create precipitation reference series, two of which involved an arithmetic mean. The first was an arithmetic mean of the homogeneous and complete stations; the second method was an arithmetic mean of normalized data so stations that were not serially complete could be used.

Young (1993) used three different techniques to objectively determine the best methodology, namely multiple discriminant analysis, multiple linear regressions, and the normalized anomaly. Looking at the results on a station-by-station basis, Young (1993) chose the middle value of the three tests and applied the interpolation scheme to that station.

Another method commonly used to infer data for a particular target station is that which uses two or more neighboring stations and assigns weights to these stations based on geographical distance. Alexandersson (1986) used a weighted mean of normalized data where the weighting was based on a distance function that was determined by spatial correlation. Area averaging was found useful for analysis of climate divisions (Keim et al. 2003) or regional analyses (Knappenberger et al. 2001) in which spatial averages are of most interest.

However, inverse distance weighting was found to have large errors when applied to interpolation of single station values (Sun and Peterson 2005a). Sun and Peterson (2005a) found that a data-driven scheme is better than the conventional location-driven scheme in interpolating precipitation data, as the former approach is more likely to catch spatial discontinuities in precipitation. The inverse distance method gives more credence to stations in close geographical proximity to a target station but its application to data interpolation was also found to be poor by Eischeid et al. (1995). An important finding of Eischeid et al. (1995) was that the inclusion of more than four stations does not significantly improve the interpolation and may in fact degrade the estimate.

Several of the data adjustments using statistical techniques employ methodologies that weight the relevance of surrounding stations to interpolate a value at a target station. However, these techniques involve data-driven weighting as opposed to equal weighting of neighboring stations or weighting based on the proximity of neighboring stations to a target station.

Guttmann (2005) used spatial tests that compared a station's data against the data from neighboring stations. The spatial regression test (SRT) does not assign the largest weight to the nearest neighbor but, instead, assigns weights according to the root mean square error (RMSE) between the station of interest and each of the neighboring stations. The SRT approach has been found in a previous study (You et al. 2004) to be more accurate than the inverse distance weighting approach for the maximum air temperature and the minimum air temperature. However, both the spatial regression and inverse distance methods were found to perform relatively poorly when the weather stations are sparsely distributed (You et al. 2004).

The data-driven scheme Sun and Peterson (2005a) found to be preferable to the inverse distance weighting was called Inverse Weighting of Square of Difference (IWSD) between neighboring stations and a target station. The IWSD method applied to precipitation is a data-determined interpolation scheme that assigns more weight to the neighboring stations with precipitation or temperature values closer to the value at the target station based on the year for which both the target and neighboring stations have precipitation. When the IWSD scheme was applied to climate normals, Sun and Peterson (2005a) found this method to outperform traditional techniques. These techniques include the SRT approach of Guttman (2005), equal arithmetic weighting for neighboring stations, and inverse distance weighting.

6. METHODOLOGY OF INTERPOLATING MISSING VALUES

The purpose of this section is to comprehensively describe the technique used to interpolate the monthly values of precipitation and monthly averages of temperature for all the COOP stations used in this study. First, the IWSD weighting technique for neighboring stations used by (2005a) will be described. However, there were some modifications made to the interpolation process, so the discussion of the Sun and Peterson technique will be followed by discussion of modifications. These modifications were based on testing of variables applied to the Sun and Peterson (2005a) technique and the minimization of errors. The error testing was done on the 44 Texas USHCN stations because of the completeness these stations' time series. Because of the extensive quality control applied to the USHCN data (Karl et al. 1990), these stations can be considered quasi-homogeneous.

a. Sun and Peterson (2005a) IWSD Method

The statistical technique of Sun and Peterson (2005a) is called Inverse Weighting of Square of Difference (IWSD) and can be applied to both precipitation and temperature. IWSD is a scheme that assigns more weight to the neighboring stations with precipitation (temperature) values closer to the value at the target station based on the year for which both the target and neighboring stations have precipitation (Sun and Peterson, 2005a). The equation of interest is Eq. (1):

$$w = \frac{1}{\sum_{i=1}^{12} (P_{neigh} - P_{target})_i^2} \tag{1}$$

In Eq. (1), i represents the month in a year from January to December, *neigh* refers to a neighboring station, and. *target* refers to a target station. The weighting scheme for temperature is identical to that used for precipitation, based on data-driven correlations between a target station and neighboring stations.

Instead of monthly precipitation totals and monthly temperature averages, Sun and Peterson (2005a) are interested in monthly anomalies to climate normals for a target station. The method to estimate normals is based on the fact that monthly anomalies at any given location are similar to those in neighboring stations. The relationship for temperature anomalies is expressed as a departure, $(T - N)_{target} \approx (T - N)_{neigh}$, where N stands for a climate normal and T for a monthly temperature. Unlike temperature (Sun and Peterson 2005b), the relationship for precipitation can be described as either a departure, $(P - N)_{target} \approx (P - N)_{neigh}$, or as a ratio, $(P / N)_{target} \approx (P / N)_{neigh}$.

Sun and Peterson (2005a) sought to determine the ideal number of stations to be used in the interpolation of monthly precipitation normals at a target station. Sun and Peterson (2005b) found that the January and July numbers for TMIN are 22 and 32 and for TMAX are 18 and 23 respectively. For precipitation, it was decided that 11 neighboring stations within ~78 km of a target station was suitable. Results indicate that errors for the stations using the ratio method are slightly greater than those using the departure method. For example, the difference of error between the two methods associated with the use of COOP data from 11 neighboring COOP stations reaches about 1.0% in January and 1.7% in July (Sun and Peterson 2005a).

b. Testing Variables on Texas USHCN Stations

The creation of a methodology based on Sun and Peterson (2005a) was to be applied to several hundred COOP stations. However, some characteristics of the interpolation process needed further investigation, including the ideal number of stations for the monthly precipitation total and monthly temperature average interpolation processes. The tests were done using the 44 Texas USHCN stations to maximize the effectiveness of the interpolation process to be applied to the COOP dataset. The 44 USHCN stations in Texas were used because these stations represent the most thoroughly quality-controlled data available (Karl et al. 1990). Therefore, interpolations for a particular target station's time series can be compared to values within that time series that can be considered quasi-homogeneous. Error values represent the residuals between the interpolation for a target station's time series and the time series data itself.

This section will discuss the interpolation process for COOP monthly precipitation values. The section following the description of the precipitation interpolation process will look at the interpolation process for temperature, but only aspects of this process that differ from precipitation. Both the precipitation and temperature interpolations are based on the IWSD weighting scheme of Sun and Peterson (2005b), but with variations in the interpolations process. The USHCN stations were used as the neighboring stations for the COOP target stations because of the completeness of their records. These variations will be explained by testing done on the 44 USHCN Texas stations.

Before the interpolation process began, it was deemed suitable to test the interpolation process on the USHCN stations. The USHCN station monthly data underwent extensive quality control procedures (Karl et al. 1990) and were used as a ground-truth dataset in the testing procedures. Therefore, any interpolated value created with the varying tests mentioned in this section for a particular month and year was compared to the actual monthly value listed. The error values in this study refer to the difference between those two values and standard errors refer to the average of these differences across a station's entire time series. A standard error is defined as the magnitude of an interpolated value subtracted from its actual value for a USHCN station given a month and year.

1) Testing the effects of data availability

The research conducted on the 44 Texas USHCN stations (Fig. 6.1) shows the effects resulting from the amount of available data. In the interpolation process, there is an inverse correlation between the amount of available data and the magnitude of standard errors. Figure 6.1 concludes that having all the data in our period of interest (1900-2001) is ideal, but this can put some limitation on the number of stations available for use in this study. While using five years is not ideal, the increase in errors is significant when less than five years of data are used. Figure 6.1 shows the combined error for the 44 USHCN stations across Texas using different lengths of time for the IWSD weighting process.

The testing shown graphically in Figure 6.1 looks at different test periods in which weights were created for neighboring stations based on a slight modification of the weighting process of Sun and Peterson (2005a). Further details about these modifications will be explained below. There were ten different period lengths examined, all ending in

2001, for which weights were created for each of the neighboring stations for a particular target station. In all ten rounds of testing, these weights were used to create interpolations for the years 1900-1949.

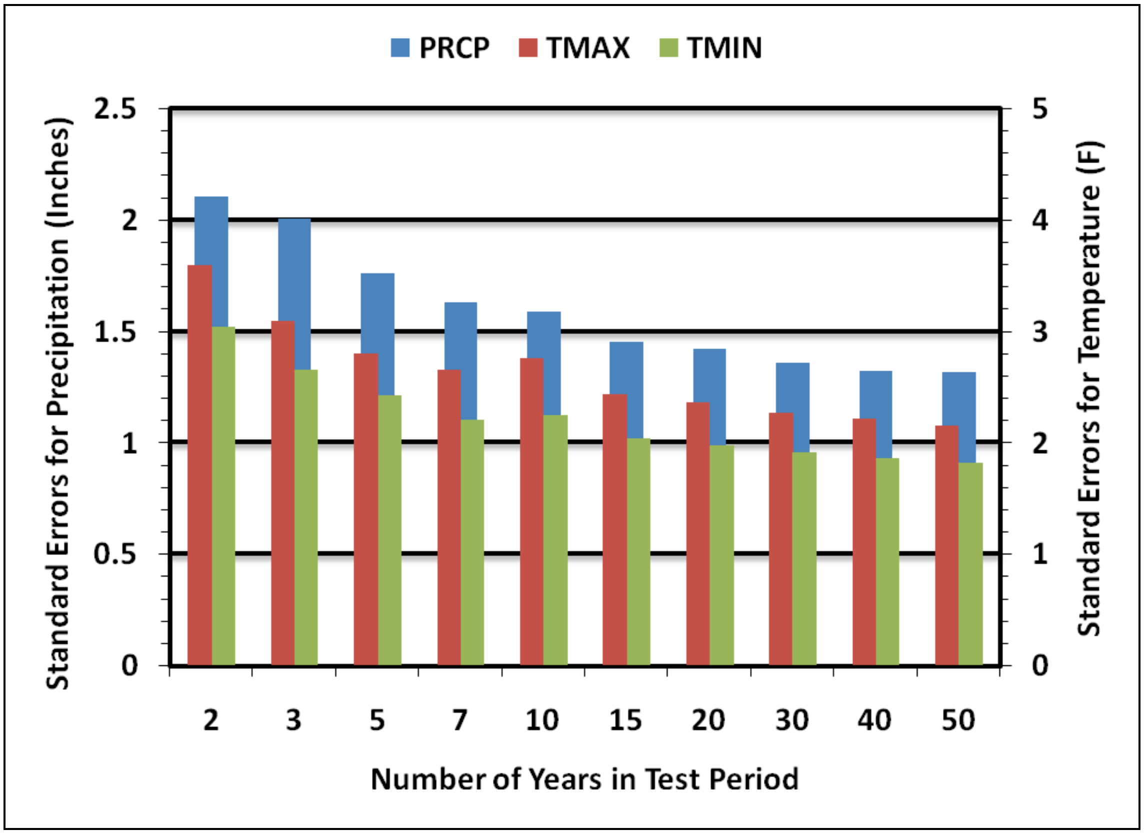


FIG 6.1. The standard errors for different test period lengths using the 44 Texas USHCN stations.

In addition to the standard errors in Figure 6.1 decreasing with increase in period length, one must consider the distributions of stations in regards to period length. Figure 6.2 shows the percentage of COOP stations with precipitation and temperature records of different lengths. One can easily see that a very small percentage of the COOP stations used in this study have period of records containing fewer than five years of data.

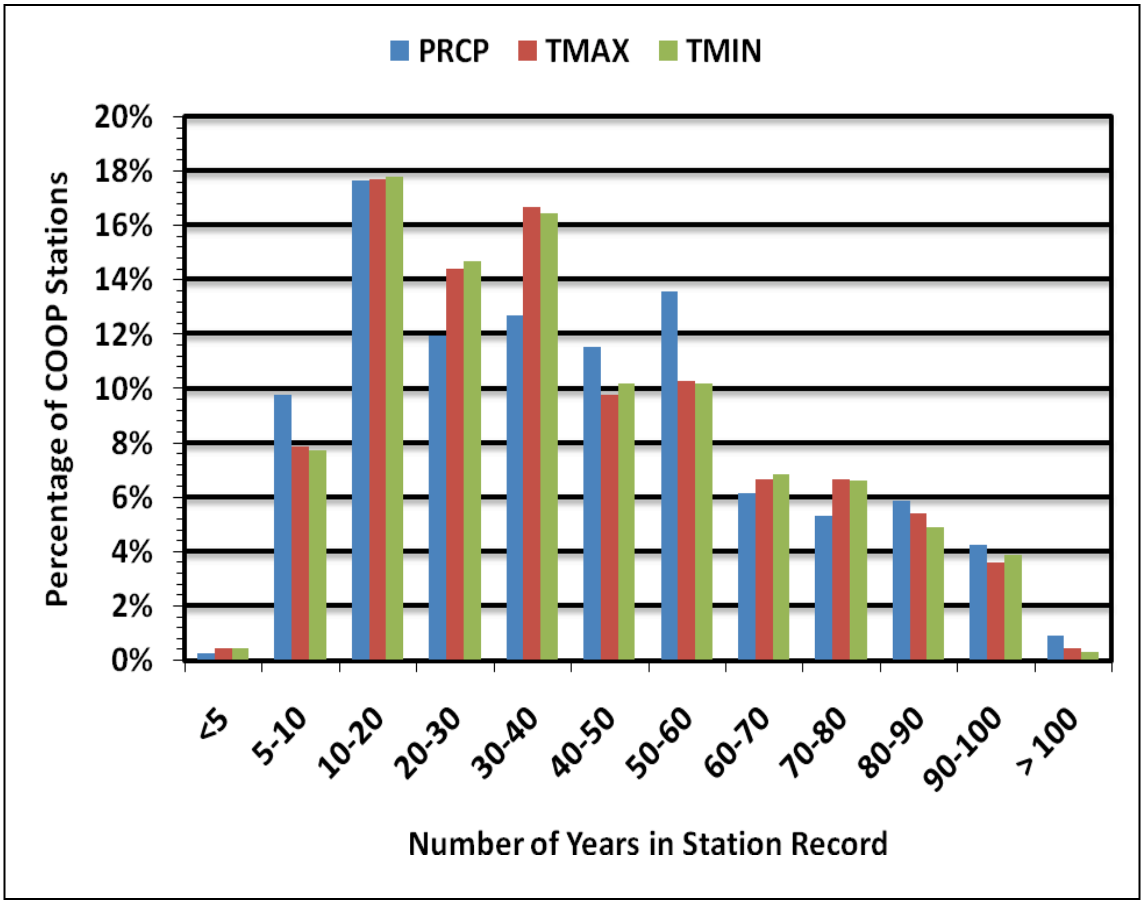


FIG 6.2. Percentage of stations with different periods of record for COOP stations in this study.

However, nearly ten percent of the COOP stations have precipitation records and eight percent have temperature records with 5-10 years of data. The extremely low percentage of stations with fewer than five years of data is the biggest factor in keeping the cutoff for COOP stations at five years rather than a larger number.

2) Testing the effects of distance

After examining the data availability issues, two different kinds of tests were run on the Texas USHCN stations testing the effects of geographical proximity of neighboring stations to the target stations. Sun and Peterson (2005a) limited the geographical proximity of neighboring stations to within 50 miles of a target station for precipitation but did not make any specifications for temperature.

Using the latitude and longitude of the COOP station, the distance from each USHCN station to the target COOP station is calculated. In order to find out if geographical distances from the USHCN stations to a specific target station were important on our interpolation process, testing was applied to the Texas USHCN stations. The standard errors averaged for the 44 USHCN Texas stations using varying geographical distance are shown in Figure 6.3.

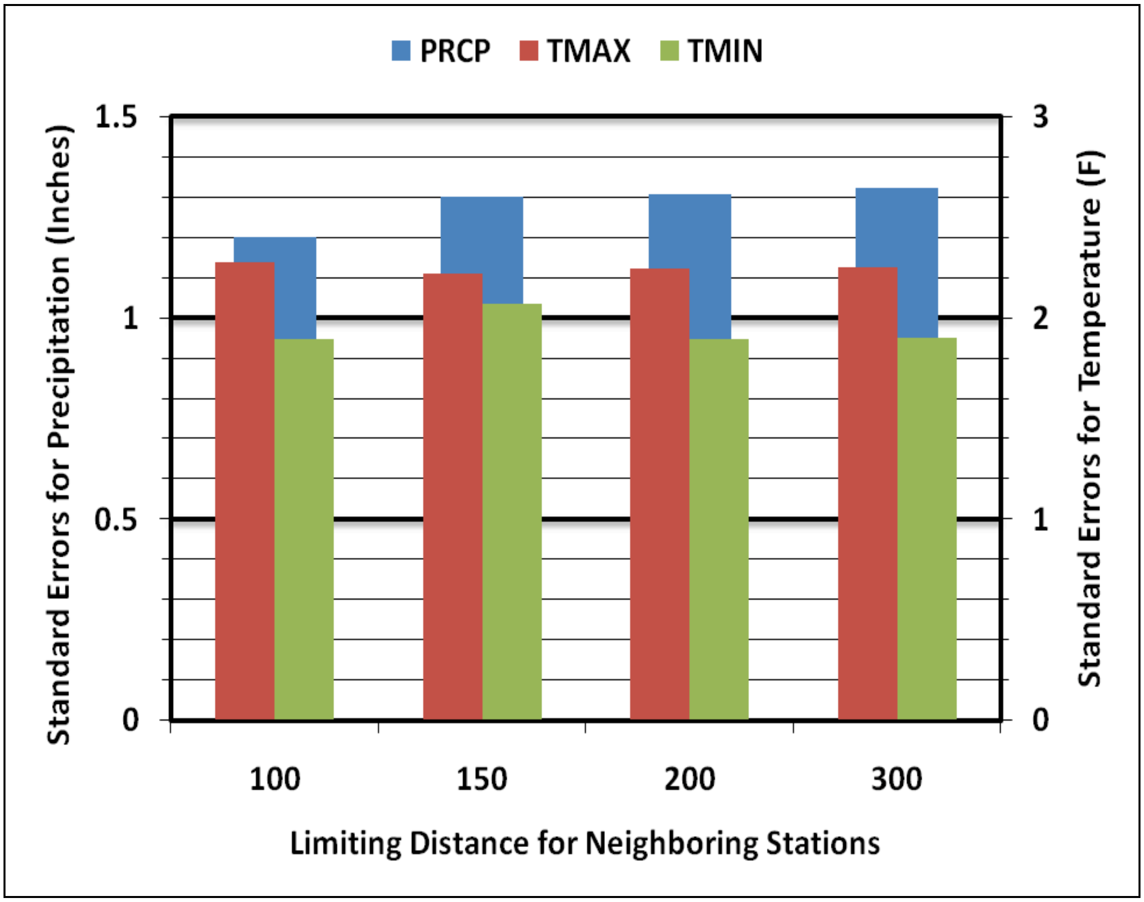


FIG 6.3. The standard errors for different distances as the limit for neighboring stations, with tests done using the 44 Texas USHCN stations.

The results shown by limiting the number of neighboring stations (Fig. 6.3) based on being within a specific radius from a target station is inconclusive for temperature. For precipitation, the standard errors are slightly lower with closer geographical proximity.

3) Testing the effects of proximity

Another testing procedure based on limiting the interpolation method to a set number of neighboring stations closest to a target station. The pattern followed by the error bars is shown in Figure 6.4 and is based on limiting a set number of neighboring stations. Figure 6.3 was based on limiting neighboring stations based on geographical distance to a target station. However, the increasing standard errors with an increasing number of stations available for interpolation are only a few percent for both precipitation and temperature.

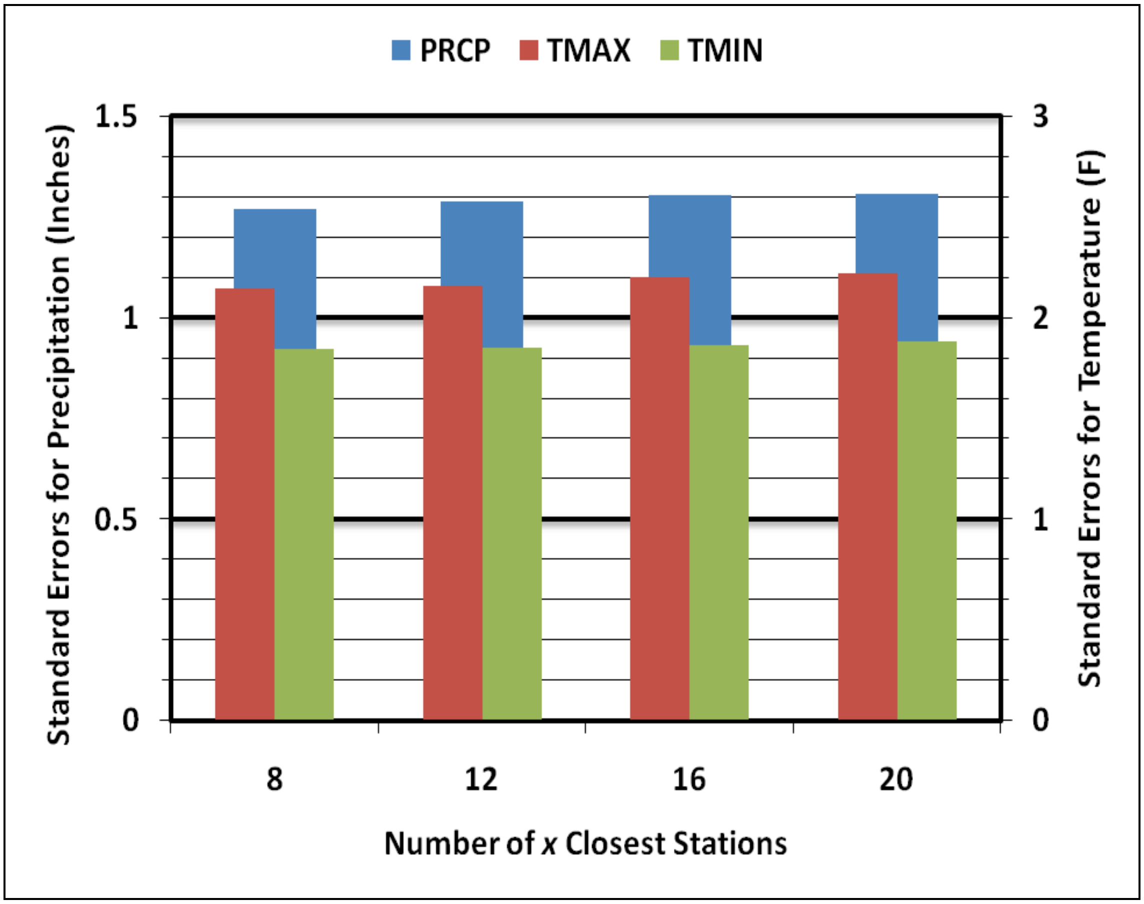


FIG 6.4. The standard errors for different distances as the limiting the *x* number of closest stations, with tests done using the 44 USHCN Texas stations.

The geographical distance test (Fig. 6.3) is inconclusive for temperature and has a slight correlation between distance and standard errors for precipitation. There is a weak correlation for all three variables in the geographical proximity test (Fig. 6.4). Because the spatial density of the USHCN network varies geographically and the results are a bit more conclusive, this study will use a limiting number of stations rather than limiting stations within a particular geographical radius of a target station.

4) Testing the effects of seasonality

After examining the effects of geography on standard errors, the effects of seasonality were tested on the USHCN stations. Eq. (1) suggests a weighting scheme for neighboring stations that includes all twelve months in one weight. Instead, this study sought to create twelve different monthly weights for each neighboring station. A month-by-month weighting scheme was preferred because of the changing spatial distributions of average monthly precipitation and temperature between months, shown graphically in the analyses of precipitation.

For a particular month and year at a COOP target station, monthly precipitation totals were compared with the corresponding monthly totals at the USHCN stations. This process is repeated for all the available months of precipitation data at a COOP target station. For each USHCN station, the differences for each of the twelve months are summed and put into a slight modification of the weighting scheme Eq. (1) used by Sun and Peterson (2005a).

The weighting scheme used in the following testing will look individually at each of the twelve months to determine which stations have the highest weight according to the available data for those months. Further investigation was done to determine the effects of using more than one month in the weighting for a particular month. This approach looks at using seasonal weights, using more than one month, as opposed to strictly monthly weights based only on correlations from the month of interest.

For a target station in this study, three different groupings of months were investigated in addition to using only data from the particular month of interest. Weighting was done using the month of interest and the previous month, the month of interest plus the following month, and the three month period centered on the month of interest. The following three graphs show the average standard error across the entire state of Texas for January (Fig. 6.5), May (Fig. 6.6), and September (Fig. 6.7).

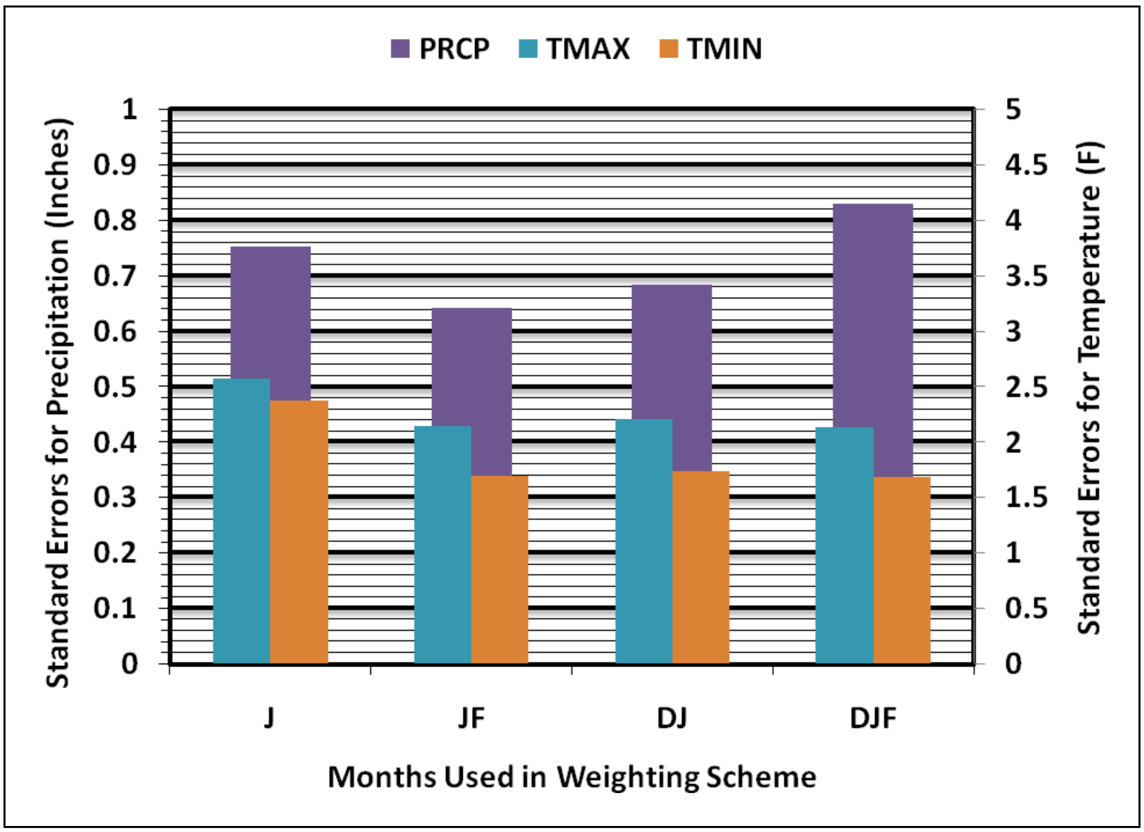


FIG 6.5. Standard errors for the 44 Texas USHCN stations using different groupings of months in the weighting of neighboring stations for the month of January in the interpolation process.

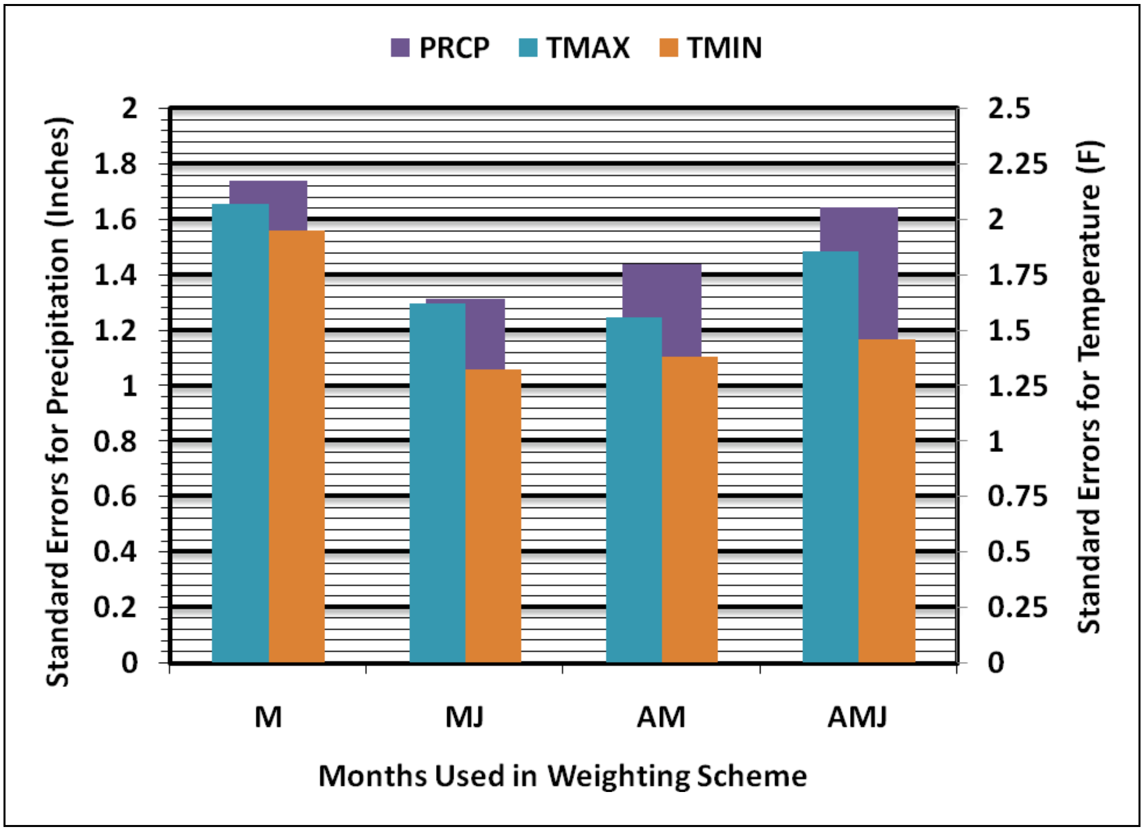


FIG 6.6. Standard errors for the 44 Texas USHCN stations using different groupings of months in the weighting of neighboring stations for the month of May in the interpolation process.

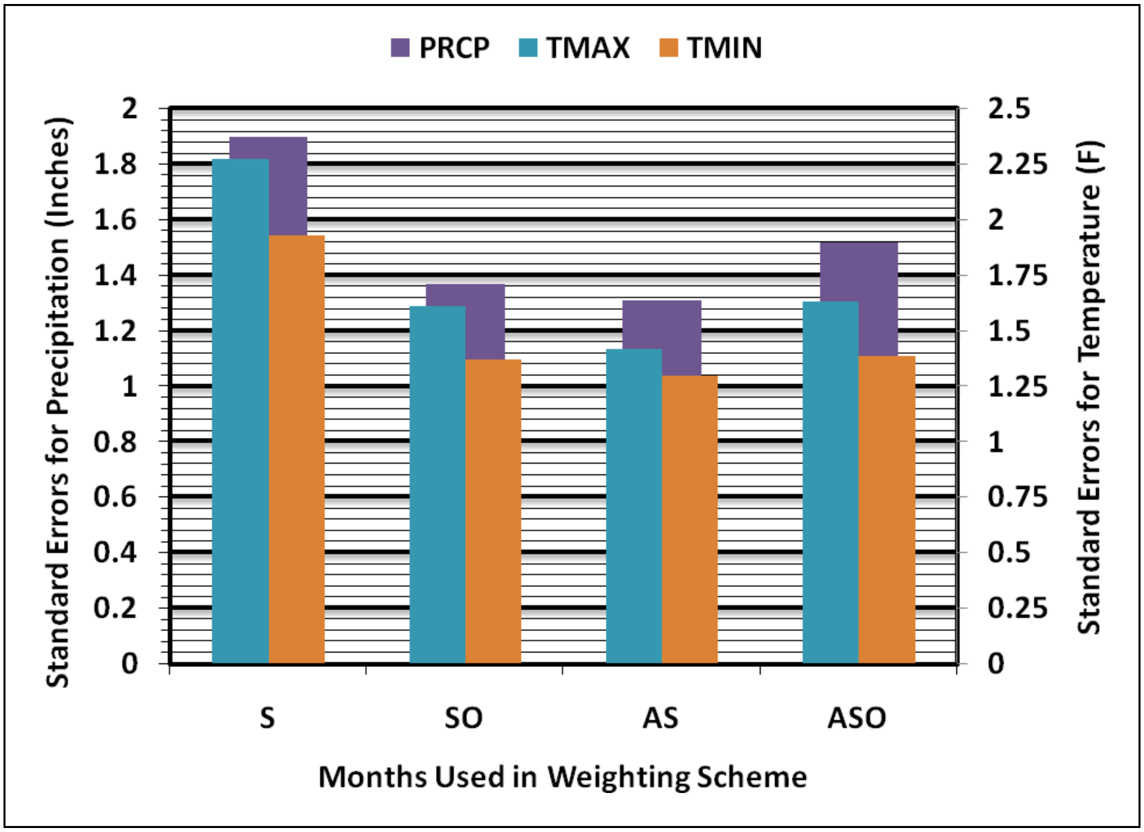


FIG 6.7. Standard errors for the 44 Texas USHCN stations using different groupings of months in the weighting of neighboring stations for the month of September in the interpolation process.

In Figure 6.6, the lowest standard errors for temperature in May averaged across the state of Texas are different for low temperatures and high temperatures. Based on Figure 6.6, one would expect low temperatures to be best represented by May and June temperatures and high temperatures best represented by April and May. The distinction is most likely spurious and does not have physical meaning.

The results showed an irregular pattern of months which provide the lowest errors for both precipitation monthly totals and temperature monthly averages across all of Texas. Because of the irregularities in the groupings of months that provided the lowest standard errors it was determined that only data from the month of interest would be used in the interpolation process, though this might increase the overall standard errors for some of the weighting.

c. Methodology Applied to Precipitation

The modified equation from Peterson and Sun (2005a) for weighting each USHCN station for a particular month is shown in the following Eq. (2).

$$Weight(neigh, month) = \frac{n}{\sum_{k=1}^{n}(value_{neigh,month,k} - value_{target,month,k})^2} \qquad (2)$$

In Eq. (2), the subscript k refers to a particular year that both the target COOP station and the USHCN station had available data and n refers to the number of years this occurred. For instance if there are 12 years in which January data appears for both the target COOP station and a USHCN station the sum will include twelve differences. For each month, the weights are normalized for all USHCN stations by multiplying by the number of years (n) in which both the COOP and USHCN data are available. This is done so that the 221 weights totaled for each of the 12 months can be accurately compared.

The total precipitation for both the COOP stations and the USHCN stations are then calculated for each month to determine a bias. Continuing on the example above, the January bias for the USHCN station would be calculated by dividing the sum of the 12 USHCN precipitation totals by the sum of the 12 COOP totals to get a ratio bias. This ratio bias is very similar to the ratio described by Sun and Peterson (2005a). The possible interpolation of negative precipitation values was the deciding factor of not using a departure bias. In addition, there were small differences in error values between the ratio and departure methods used by Sun and Peterson (2005a). A bias greater than one indicates a USHCN station has a wet bias compared to the target COOP station for that month. The bias of each USHCN station for a particular month with respect to a target station is described by the following Eq. (3).

$$Bias(neigh, month) = \frac{\sum_{k=1}^{n} value_{neigh,\,month,\,k}}{\sum_{k=1}^{n} value_{target,\,month,\,k}} \tag{3}$$

The subscript k in Eq. (3) refers to a particular year in which both data from the COOP and target USHCN station were available for the month of interest and n refers to the number of years this occurred. For both the weighting and bias calculations for the twelve months at each USHCN station, less randomness will occur with an increase in the available number of data.

The final few steps of the interpolation involve a slight modification of the Sun and Peterson (2005a) IWSD scheme in which distance becomes an important variable. Starting with January 1900 through December 2001, each month is analyzed to see if an interpolated value can be created, which leads to a serially complete record for PRCP, TMAX, and TMIN.

For each month in this period, the closest twenty USHCN stations are analyzed to see if four or more stations have data available for this month. If so, the four stations with the highest weights for this particular month of the year are used to create an interpolated value. The use of four stations in the final interpolation value is based on the work of Eischeid et al. (1995), which concluded using four target stations was ideal using this type of interpolation scheme.

The following Eq. (4) shows how this value is calculated for any particular month in this period.

$$Interpolated\ value\,(month,\ year) = \frac{\sum_{i=1}^{4} weight_{i,month} \times value_{i,month,year}}{\sum_{i=1}^{4} weight_{i,month} \times bias_{i,month}} \qquad (4)$$

When there are fewer than four stations available for a month and year within the twenty closest USHCN stations to the target station, the process is repeated continuously by adding the next closest station until four stations with data for a particular month and year are found. For instance, if the 20 closest USHCN stations to a target station for January 1900 yield only three with data available, the program will continue to look for the next closest station until one with data is found. This process keeps distance as an important variable but assures that four stations will be used in the interpolation regardless of their distance to the target station.

d. Methodology Applied to Temperature

There are both similarities and differences in the interpolation of temperature monthly averages compared to the process described for precipitation monthly averages. The COOP target station data is analyzed to determine the mean temperatures for all twelve months, and then each target station monthly average is given an anomaly according to its mean for that month. However, the mean is calculated only for months in which both actual data from the target station and data for the USHCN neighboring station being weighted is available. The following Eq. (5) describes the calculation of the anomaly at each available COOP data point.

$$Anomaly\,(Month, year) = value_{month,\,year} - mean_{month} \tag{5}$$

Likewise, twelve monthly means are calculated for each of the 221 USHCN stations over their entire time series. For each USHCN station, these twelve means are used to calculate the anomalies for each monthly average in its period of record using the above equation. The USHCN weights for temperature are based on the differences in anomalies, whereas the precipitation USHCN weights were based on differences in the actual monthly precipitation totals.

Differences were calculated for the temperature data almost exactly as with the precipitation data, the exception being that monthly anomalies replaced the monthly averages. The USHCN weights were calculated using Eq. (6) for a particular month of the year.

$$Weight\,(USHCN\ station, month) = \frac{n}{\sum_{k=1}^{n}(anomaly_{USHCN,month,k} - anomaly_{target,month,k})^2} \qquad (6)$$

For each month in the time period, the interpolated values were calculated using the four highest weighted stations as in the interpolation process for precipitation. The use of anomalies eliminated the need for biases in the final interpolation. The following Eq. (7) describes the interpolation of an anomaly for a particular month and year in the time period between January 1890 and December 2001.

$$Interpolated\ anomaly(month,\ year) = \frac{\sum_{i=1}^{4} weight_{i,month} \times anomaly_{i,month,year}}{\sum_{i=1}^{4} weight_{i,month}} \qquad (7)$$

The interpolated anomaly was then added to the COOP target station mean value throughout its time series to get a final interpolated value for each month and year, shown in Eq. (8).

$$Interpolated\ value\ (month,\ year) = Interpolated\ anomaly\ (month,\ year) + Mean\ value\ (month) \qquad (8)$$

7. AVAILABLE DATASETS

a. COOP and USHCN Datasets

The analysis of precipitation trends in relation to drought across New Mexico and Texas will take advantage of several datasets at our disposal, subsets of the COOP and the USHCN datasets. Again, though the quality control and interpolation procedures for temperature have been thoroughly discussed, analyses of these datasets will not be included. The USHCN dataset is highly quality-controlled with few gaps in stations' time series but with poor spatial coverage. Figure 7.1 is a bar graph that examines the number of years of available data for the 221 USHCN stations used in this study.

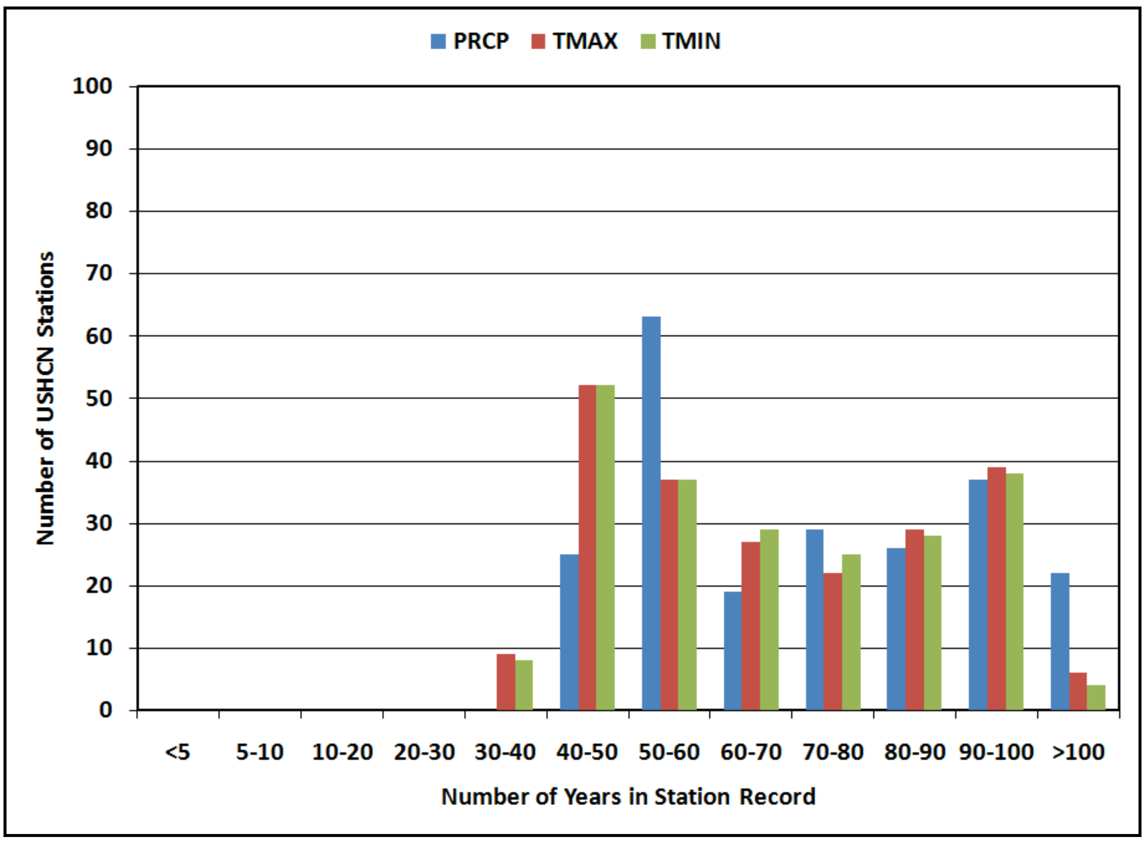


FIG 7.1. Bar graph of different groupings for the number of years of available data for the USHCN stations in this study. This graph represents all the USHCN stations in this study and shows percentages for PRCP, TMAX, and TMIN.

More than 95% of the USHCN stations in Arizona, Arkansas, Colorado, Louisiana, New Mexico, Oklahoma, Texas, and Utah have at least 40 years of data available. The main strength of the COOP dataset is its spatial coverage but a large percentage of the stations have either short periods of records or large gaps of missing data. Figure 5.2 showed a bar graph displaying the percentage of COOP stations having different groupings of data availability for the 33 climate divisions used in this study. This graph is based on all the COOP stations before the quality control check was applied since the COOP stations with less than five years of data were eliminated prior to the quality control procedures.

b. Groupings of USHCN Stations

The USHCN dataset used in this study was subdivided into two sections, the first containing the ninety-six stations with periods of record lasting through the entire 20^{th} century (Fig. 7.2). The second subset contains the stations the rest of the USHCN stations, many of which have a period of record beginning in 1948. Only 2 of the 44 USHCN stations in Texas have data starting in 1948, whereas 17 of the 24 New Mexico USHCN stations station records began in 1948. Analyses done on the USHCN data will focus mainly on the actual data from the long-term stations and the interpolated data sets using all 221 USHCN stations used in this study.

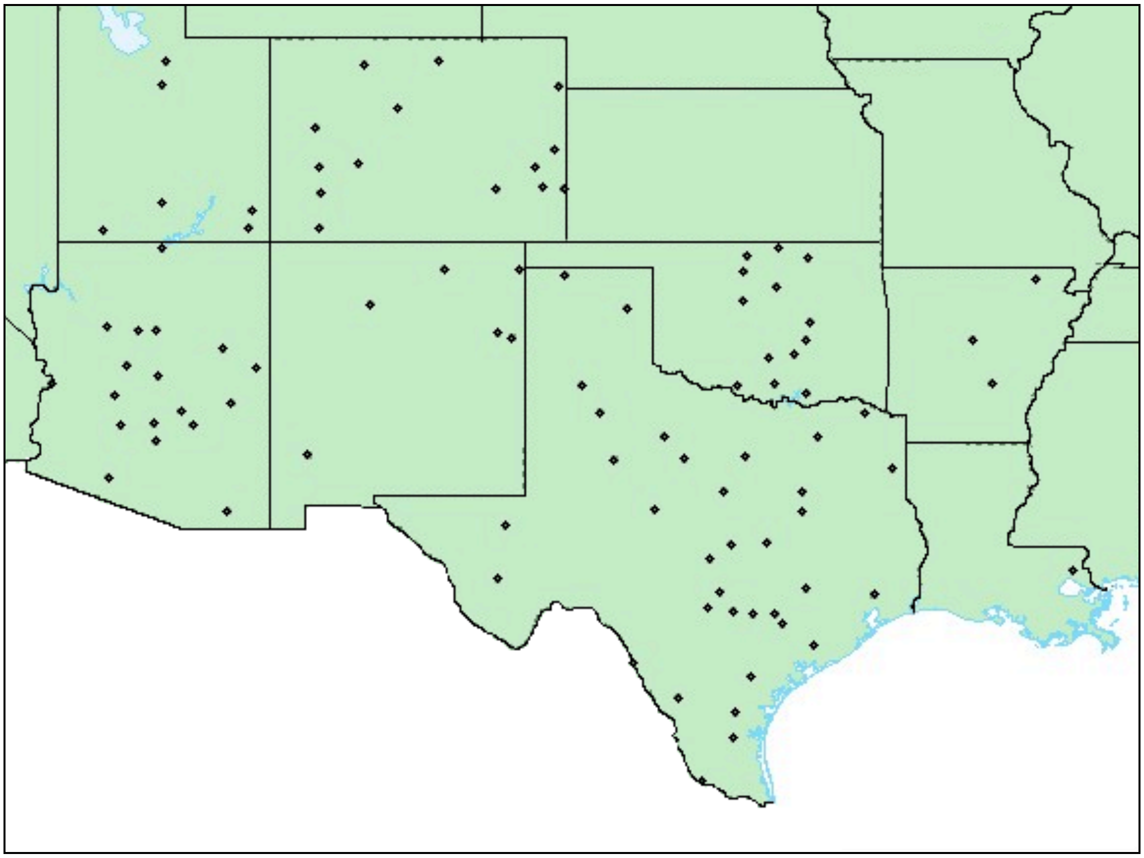


FIG 7.2. Map of the 96 long-term USHCN stations.

Another grouping of the USHCN stations splits them into homogeneous and inhomogeneous stations based on analyses performed on station metadata. A detailed, subjective analysis of the New Mexico and Texas USHCN stations was done and was based on station relocations. The stations in the USHCN dataset have long-term periods of record, but most stations have moved from their original location to one or more locations within a small distance. Using the station history coordinates, elevations, and descriptions, station moves were investigated to uncover discontinuities in station records.

A table summarizing a detailed investigation into the station histories for the 44 Texas and 24 New Mexico USHCN stations is found in Appendix B. A detailed methodology that describes the criteria and process for finding the stations not deemed completely homogeneous is found following the summary table in Appendix C. About half of the USHCN stations in New Mexico and Texas were deemed to be homogeneous stations.

c. Groupings of COOP Stations

The major groupings of the COOP data are by climate division. Figure 7.3 is a United States climate division map with the divisions of interest shaded and labeled according to its numbering within its own state. The climate divisions of interest are all those in New Mexico and Texas, as well as divisions bordering either Texas or New Mexico. Several of the time series analyses in Section 8 are grouped together and represent the New Mexico (Fig. 7.3, blue), West Texas (Fig. 7.3, green), and East Texas (Fig. 7.3, gray) regions.

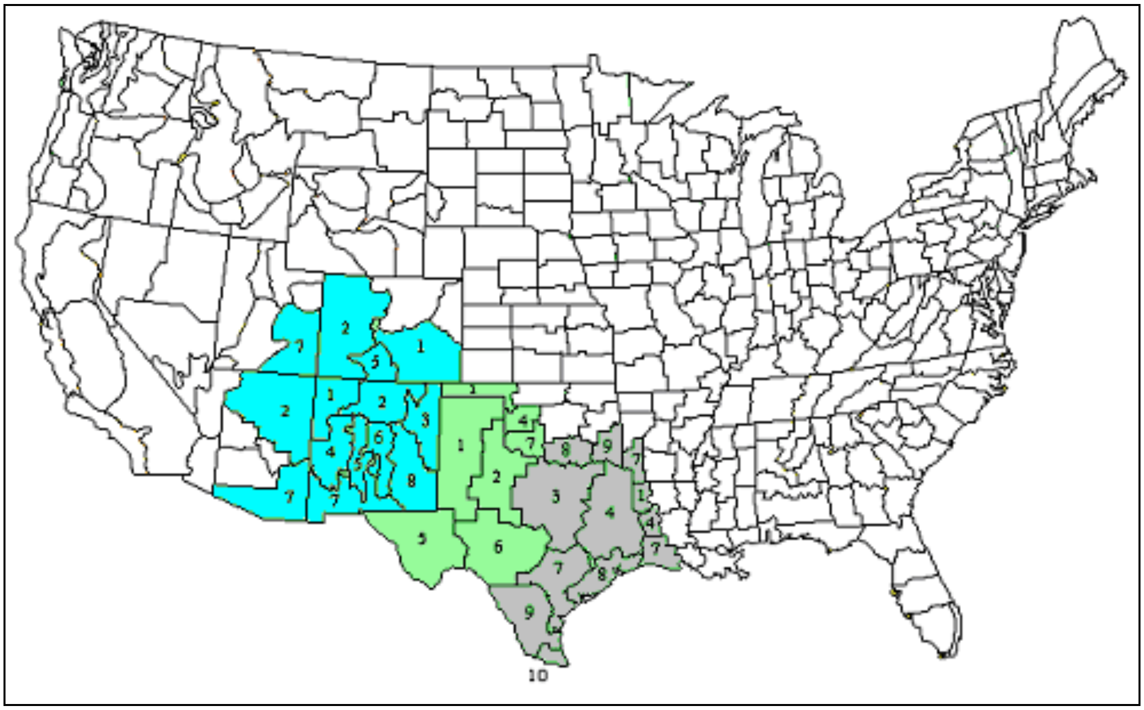


FIG 7.3. United States climate division map with divisions of interest shaded and numbered. The blue shaded climate divisions represent the New Mexico region, the green represent the West Texas region, and the gray represent the East Texas region in this study.

Additionally, the COOP stations underwent the extensive quality control procedure described earlier in which outliers and unrealistic data were eliminated. The remaining data for all the COOP stations went into a subset of data called “Quality Control Data.” Before this quality control was performed on the COOP station data, a superset of the data called “Raw Data” was created to preserve the data eliminated in the quality control checks. The “Raw Data” subset includes the passing and failing the quality control checks. However, the analyses done on the COOP data in this study will contain only those done on the Quality Control data subset.

d. Types of Data in the COOP and USHCN Datasets

The original data obtained from NCDC in both the COOP and USHCN datasets were daily precipitation totals, daily maximum temperatures, and daily minimum temperatures. The primary intent of the COOP program is the recording of 24-hour precipitation amounts, but about 55% of the stations also record maximum and minimum temperatures (NCDC 2006). For this reason, the daily precipitation data are more numerous than the daily temperature extremes data.

Data on monthly time scales rather than daily time scales were more desirable for the purposes of this study, so the daily values were transformed to monthly values. For precipitation, the daily values at each USHCN and COOP station were summed into a monthly precipitation total. For both maximum and minimum temperature, the daily totals were averaged across the entire month to create a time series of monthly averages at each USHCN and COOP stations.

In addition to time series of monthly precipitation totals, multiple-month precipitation accumulations were created for each USHCN and COOP station. The specific time periods in which accumulations were calculated were 3, 6, 12, 24, and 48 months. These accumulations at each station are a time series of precipitation accumulations for a given length of time up to a particular month. For example, a 3-month accumulation precipitation total for August 2001 would include the June through August precipitation total for a given station.

Similarly, multiple-month averages of both maximum and minimum temperature were calculated over the same 3, 6, 12, 24, and 48-month periods in which data were available. For both precipitation and temperature, these calculations were done only when every month in the period contained available data.

Decadal averages for the USHCN and COOP data were calculated both on monthly and annual timescales for the precipitation and temperature records at each station. For instance, the February average maximum temperature at each COOP and USHCN station was calculated using the available February averages for the time period 1901-1910, 1911-1920, and etc. These decadal statistics also exist for the annual precipitation totals and annual maximum and minimum temperature averages.

e. Derivation of Precipitation Values

For each type of precipitation data described, there were three different ways to derive data for a given month and year at a particular USHCN or COOP station. The first way is done by simply assuming the monthly precipitation total given for the USHCN or COOP station. It is very important to note that other derivations of precipitation values are used only when actual data for a station is not available.

The second derivation uses the interpolation scheme discussed previously and derived from Sun and Peterson (2005a). Four USHCN stations are used to derive an interpolated value for both the USHCN and COOP stations. This interpolated value is calculated whenever four USHCN stations are available and is not dependent on the availability of an actual value. Even if an actual USHCN or COOP data value is available, an interpolated value is still calculated. Because of the completeness of most USHCN precipitation time series, these interpolated values are serially complete for the period from roughly 1900-2001. These values are used when actual precipitation values are not available for a given month.

The third derivation monthly precipitation totals for a given USHCN or COOP station is based on gamma distribution statistics. Research on the fit of gamma distributions to monthly precipitation distributions stretches back several decades. Barger and Thom (1949) found that the gamma distribution provided good fit to precipitation series in the United States. Momiyama and Mitsudera (1952) fit the gamma distribution to monthly rainfall totals over Japan while Mooley (1972) used a gamma distribution modal for the Asian summer monsoon. Klein and Bloom (1987) found filling monthly precipitation totals using a gamma distribution was desirable. The NCDC bases their monthly precipitation probabilities from the 1971-2000 United States Climate Normals on fitting a gamma distribution.

The actual data time series and interpolated data time series for a station can be represented as distinct gamma distributions. Eq. (9) describes the probability of a particular monthly total *x* for a given series. The gamma distribution is a two-parameter family of continuous probability distributions with shape parameter α and scale parameter β

$$f(x,\alpha,\beta) = x^{\alpha-1}\,\frac{e^{-x/\beta}}{\beta^{\alpha}\Gamma(\alpha)} \text{ for } x > 0 \text{ and } \alpha, \beta > 0 \qquad (9)$$

Because the data is limited for each time series, it is necessary to estimate the shape parameter $\hat{\alpha}$ and the scale parameter $\hat{\beta}$ using another quantity *D* (Wilks 2006). The equation for *D* in Eq. (10) is given and uses the mean of the time series and the natural logarithms of each value within the time series. Using this quantity D for a given time series, whether it is an actual time series or interpolated time series, it is possible to find the estimator $\hat{\alpha}$ given in Eq. (11). After the estimator $\hat{\alpha}$ is found, it is then possible to find the estimator for the size parameter $\hat{\beta}$ using Eq. (12) given $\hat{\alpha}$ and the mean of the time series $\overline{X}$. The gamma distributions for the actual and interpolated time series for a given station determines the sample mean $\overline{X}$ and sample variance s^2, shown in Eq. (13) for each series.

$$D = ln\left(\sum_{i=0}^{n} x_n\right) - \sum_{i=0}^{n} ln(x_i) \tag{10}$$

$$\hat{\alpha} = \frac{1 + \sqrt{1 + (4D/3)}}{4D} \tag{11}$$

$$\hat{\beta} = \frac{\overline{X}}{\hat{\alpha}} \tag{12}$$

$$s^2 = \alpha\beta^2 \tag{13}$$

In an ideal world the sample mean of the actual time series and interpolated time series for a particular station would be equal. However, the interpolated time series generally samples a larger time period whereas most stations have an actual time series with large gaps in the record. For instance, if a COOP station has actual values (sample population) only taken in a relatively wet period, the actual mean for that distribution will be larger than the interpolated mean for that same station.

Therefore, the interpolated time series sample mean ($\overline{X}_{interpolated}$) and actual time series sample mean ($\overline{X}_{actual}$) are unequal for the vast majority of COOP and

USHCN stations. Because $\overline{X}_{interpolated}$ is in the majority of cases based on a more temporally complete time series and larger sample population, the mean of the variance-adjusted time series ($\mu_{variance-adjusted}$) is assumed to be equal to $\overline{X}_{interpolated}$.

Using the estimators of the shape $\hat{\alpha}$ and scale parameter $\hat{\beta}$ for each series at a particular station, cumulative distribution function (CDF) values for each series were calculated. The CDF value for a given random variable X represents the probability that the X takes on a value less than or equal to x, where $0 \leq x \leq 1$ (Wilks 2006). Precipitation values were increased incrementally from zero by a hundredth of an inch, with each value assigned a CDF value. This process is repeated until the CDF value reaches one. For each hundredth of an inch, the CDF value represents the probability of precipitation being less than or equal to that monthly precipitation total.

A large percentage of the COOP time series have periods of records shorter than 20 years (Fig. 5.2). This small sample population might lead to unrealistic means and variances for these stations. The variances of all the interpolated precipitation time series are unrealistically low. In order to get a variance representative of a century-long time series, the variance of this third series ($\sigma^2_{variance-adjusted}$) was assumed to be an average of the two closest USHCN station variances. The two USHCN station variances were based on the actual time series data and not the interpolated time series of values. Therefore, $\sigma^2_{variance-adjusted}$ is more indicative of the variance of actual time series data for a given station.

The spatial nature of the variance distributions is discussed in the analysis section but the interpolated dataset variances for each station are generally smaller than the actual dataset variances. Because each interpolated value is a weighted average of four stations, extreme actual values may be moved closer to the mean value for a data series by the interpolation if the actual monthly value is isolated in nature. This can be especially problematic of summertime precipitation which is very erratic and can produce isolated monthly rainfall totals far exceeding the mean total of surrounding regions.

For a specific station, the data in the third time series uses data from the actual time series and interpolated time series for that station. In the case an interpolated monthly value is available, the actual value is assigned a CDF value based on the distribution of interpolated precipitation values for that station. The third time series precipitation value for that particular month is the precipitation value that matches the CDF value in the variance-adjusted distribution. For months an actual value is available, that actual value is used in the time series. Therefore, the variance-adjustment procedure is much like the interpolation process in that it is used to fill missing values in the actual data time series for a station.

The monthly values in this third variance-adjusted time series are based on CDF values from individual months but with an overall distribution characteristic of the homogeneous USHCN stations in close proximity. However, the sample mean $\overline{X}_{variance-adjusted}$ of the overall third time series may differ slightly from $\mu_{variance-adjusted}$.

8. BASIC ANALYSIS OF PRECIPITATION DATASETS

a. Naming Convention of Datasets

The analyses of the datasets will be divided into those done on the COOP datasets and those done on the USHCN data. The analyses done on the COOP data will consist of climate division averaging due to the abundance of stations. The analyses of the USHCN data will be done on a station-by-station basis because this dataset can be managed more easily. Time series graphs showing temporal changes in variables will use the USHCN dataset for statewide averages and the COOP dataset for climate division averages. Analyses will also be related to drought and extreme precipitation in New Mexico and Texas.

However, with the large number of datasets available, the use of a system to abbreviate the names of the different datasets will come in handy. The two main datasets are the USHCN and the COOP datasets, with each COOP dataset containing a quality-controlled subset and a “Raw Data” superset for which data have not been flagged and eliminated by quality control processes. Table 8.1 lists the different dimensions of the COOP and USHCN datasets and the single-characters that will be assigned to variables in each dimension.

An important note is that each dataset contains the actual data from a COOP or USHCN station when available. Each dataset differs in that it uses a different method to fill in the missing data in the time series of a station. Therefore, a large percentage of the precipitation monthly values in each dataset are exactly the same. The values differ for a given month or time period only when the actual data is not available. However, several of the COOP stations have short periods of record, so the procedure of filling the missing values is particularly important for these stations.

TABLE 8.1. The single-letter abbreviation for the different dimensions of the COOP and USHCN datasets. Each data set is represented as a three-letter permutation of these dimensions.

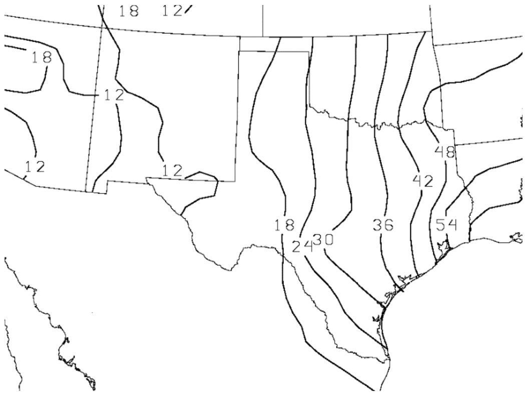
1. Coop Data (C)	**1. USHCN Data (U)**
2. Time Series Type	**2. Time Series Type**
• Actual (A) • Interpolated (I) using all USHCN stations • Interpolated using high-quality USHCN stations (Q) • Interpolated and • variance-adjusted (V) • Interpolated and variance-adjusted using high-quality USHCN stations (q) • Interpolated, variance-adjusted, and trend-adjusted using high-quality USHCN stations (T)	• Actual (A) • Actual using only long-term USHCN stations (L) • Trend-adjusted using only actual data from long-term USHCN stations (T) • Interpolated (I) using all USHCN stations • Interpolated and variance-adjusted (V)
3. Data Period	**3. Data Period**
• Monthly (M) • Annual (Y)	• Monthly (M) • Annual (Y)

Every dataset used in the analyses of this section can be written as a permutation of the three dimensions that are included on Table 8.1, beginning with the dataset as either COOP (C) or USHCN (U). For instance, when an analysis uses the COOP actual annual precipitation dataset, the code for this dataset is CAY. It should be mentioned that the study done to determine if USHCN stations were high-quality, summarized in Appendix B was done only in New Mexico and Texas. Otherwise, the COOP datasets contain the 18 climate division in New Mexico and Texas and the 15 that border these two states.

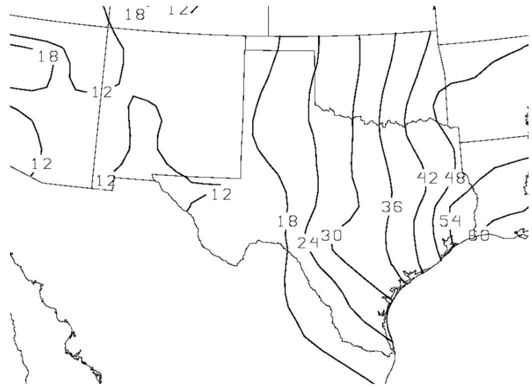
Additionally, it must be emphasized that the analyses are based solely on the "Quality Control" subset of data contained within each dataset. The "Raw" dataset is available to be analyzed but was not done so in this particular study.

b. Annual, Seasonal, and Monthly Average Precipitation

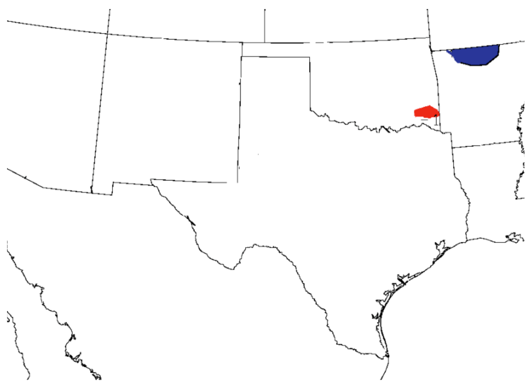
The analyses of the datasets will start with precipitation, and the most basic analysis could be considered the annual averages for USHCN stations using the ULY dataset, which shows the spatial differences in average but not the temporal differences in the three different time series for individual stations. Figure 8.1 contains the annual averages for precipitation using the UIY and UAY datasets and the difference between these two datasets. Though using actual data from USHCN long-term stations (ULY) would likely give the truest mean value at any location, Figure 7.2 shows that the spatial distribution of these stations to be sparse, especially in New Mexico.



(a) UIY



(b) UAY



(c) Difference between UIY and UAY (UIY-UAY)

FIG 8.1. Average annual precipitation (inches) at USHCN stations for the UIY dataset (a) the UAY dataset (b) and the difference (c) between the two datasets. A difference greater than one inch (UIY > UAY) is denoted by the red shading and blue shading (UAY > UIY).

Though there are slight differences between the maps on Figure 8.1, these differences are almost impossible to see without Fig 8.1c. Moving from east to west, annual precipitation values decrease at a relatively uniform rate with the gradient decreasing slightly in West Texas. An absolute minimum in the geographical domain of this study occurs in the eastern region of New Mexico CD 8, the Southern Desert region of New Mexico. The maxima in annual average precipitation occur in the southeastern corner of our domain in Louisiana approaching 60” as an annual average. The COOP climate division averages break down the annual averages into specific numbers (Fig. 8.1).

The three monthly precipitation distributions (actual, interpolated, and variance-adjusted) were averaged for each month using the available stations within the COOP climate divisions. Three new monthly distributions for each climate division in the geographical domain of this study were formed to be used for climate division statistics. Table 8.1 shows the annual average precipitation total for each COOP climate division in this study for the CAY, CIY, and CVY datasets.

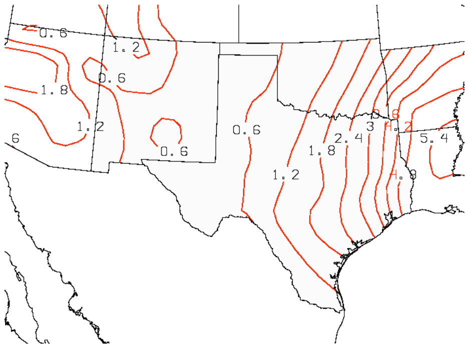
TABLE 8.2. Annual average precipitation for COOP climate divisions.

Climate Division	CAY	CIY	CVY
Arizona 2	14.08	15.12	14.63
Arizona 7	14.25	15.24	15.26
Arkansas 7	50.61	52.37	52.50
Colorado 1	14.96	15.20	15.19
Colorado 2	16.56	17.52	17.35
Colorado 5	11.10	11.69	11.70
Louisiana 1	49.41	49.80	49.79
Louisiana 4	52.85	53.79	53.77
Louisiana 7	59.20	58.87	58.83
New Mexico 1	10.85	11.16	11.07
New Mexico 2	16.62	16.24	16.14
New Mexico 3	15.77	15.98	15.93
New Mexico 4	14.42	14.19	14.24
New Mexico 5	9.45	10.00	9.46
New Mexico 6	16.22	17.63	17.56
New Mexico 7	13.80	13.62	13.64
New Mexico 8	11.00	11.70	11.60
Oklahoma 1	22.92	25.71	26.13
Oklahoma 4	26.80	26.37	26.36
Oklahoma 7	28.25	28.41	28.42
Oklahoma 8	37.07	38.00	38.02
Oklahoma 9	47.40	48.62	48.25
Texas 1	19.52	18.92	18.93
Texas 2	23.47	23.59	23.59
Texas 3	33.68	33.82	33.84
Texas 4	45.27	46.02	46.02
Texas 5	12.76	12.61	12.51
Texas 6	25.75	25.15	25.09
Texas 7	33.54	33.57	33.57
Texas 8	45.98	46.51	46.50
Texas 9	23.25	23.60	23.40
Texas 10	24.26	23.86	23.83
Utah 7	9.61	9.70	9.77

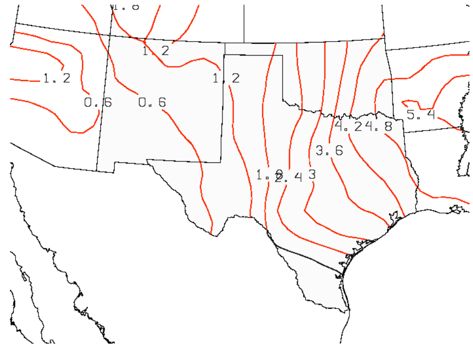
The climate division COOP dataset annual average values are in good agreement with the USHCN dataset. However, within climate divisions, the mean annual precipitation is different among the three different datasets. Again, this would be due to the sample population, especially since $\mu_{variance-adjusted}$ is assumed to be equal to $\overline{X}_{int\,erpolated}$.

As one would expect, the largest annual values are in the climate divisions on the eastern boundary of our domain in Arkansas and Louisiana. An interesting climate division is Texas CD 8, with an annual average precipitation around 45 inches per year. However, the USHCN map shows this climate division to have the strongest spatial gradient across its climate division, with the areas near the Texas-Louisiana border in Texas CD 8 averaging between 55-60 inches of precipitation per year. The driest overall conditions are in the Central Valley CD 5 in New Mexico and the Southeast CD 7 in Utah. New Mexico is extremely dry compared to the whole of Texas, but it appears that Texas CDs 1 and 5 can be grouped with the New Mexico climate divisions based on annual average precipitation.

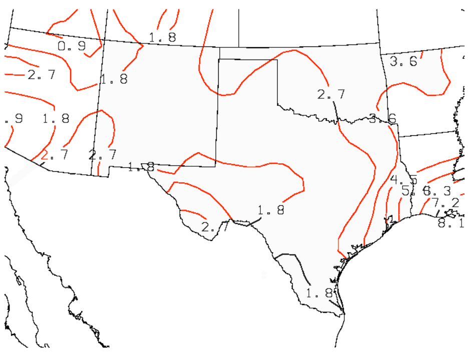
The spatial distribution of average annual precipitation is not necessarily indicative of precipitation in individual seasons. The monthly averages were plotted using the UIY dataset for January (Fig. 8.2a), April (Fig. 8.2b), July (Fig. 8.2c), and October (Fig. 8.2d) to show both the differences in the spatial gradient of precipitation between seasons and the changes in mean precipitation throughout the year at individual stations. Additionally, the monthly averages calculations using the CVM dataset are summarized in Table 8.2.



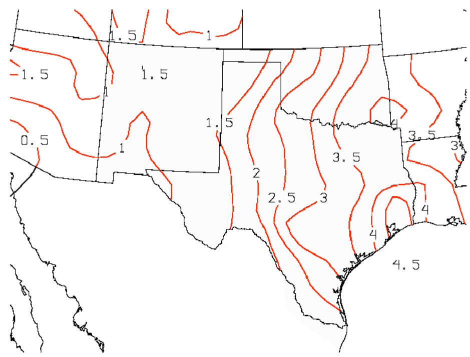
(a) January



(b) April



(a) July



(b) October

FIG 8.2. Average January (a), April (b), July (c), and October (d) precipitation at USHCN stations (UIY dataset).

TABLE 8.3. Monthly average precipitation for COOP climate divisions (CVM).

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Arizona 2	1.29	1.28	1.38	0.90	0.58	0.46	2.03	2.33	1.47	1.24	1.05	1.30
Arizona 7	1.10	1.07	0.95	0.47	0.28	0.53	3.09	3.05	1.60	1.02	0.87	1.30
Arkansas 7	4.09	3.86	4.97	5.40	5.77	4.33	3.96	3.25	3.71	3.89	4.54	4.76
Colorado 1	0.46	0.55	1.03	1.52	2.09	1.63	2.47	2.22	1.19	0.93	0.65	0.53
Colorado 2	1.37	1.35	1.62	1.61	1.45	1.05	1.65	1.84	1.60	1.52	1.26	1.34
Colorado 5	0.55	0.57	0.82	0.92	1.01	0.77	1.83	1.97	1.20	1.03	0.60	0.61
Louisiana 1	4.74	4.30	4.50	4.89	4.99	3.96	3.95	3.05	3.13	3.31	4.36	5.07
Louisiana 4	5.30	4.35	4.75	4.94	5.35	4.15	4.69	3.63	3.39	3.17	4.66	5.77
Louisiana 7	5.19	4.11	4.33	4.50	5.29	5.09	6.86	5.36	4.41	3.76	4.50	5.78
New Mexico 1	0.74	0.74	0.88	0.71	0.62	0.53	1.63	1.88	1.26	1.03	0.74	0.77
New Mexico 2	0.73	0.76	1.07	1.15	1.50	1.37	2.60	2.84	1.66	1.27	0.82	0.76
New Mexico 3	0.33	0.39	0.64	1.07	2.05	2.02	2.77	2.78	1.80	1.25	0.57	0.46
New Mexico 4	0.75	0.72	0.76	0.55	0.58	0.76	2.60	3.00	1.80	1.18	0.68	0.91
New Mexico 5	0.36	0.38	0.46	0.37	0.45	0.62	2.85	2.42	1.63	1.08	0.45	0.52
New Mexico 6	0.87	0.95	1.04	0.90	1.16	1.35	2.84	2.98	1.90	3.12	0.87	1.08
New Mexico 7	0.36	0.40	0.47	0.69	1.38	1.55	2.24	2.36	2.12	1.35	0.51	0.49
New Mexico 8	0.66	0.66	0.52	0.28	0.35	0.60	2.26	2.37	1.64	1.02	0.61	0.88
Oklahoma 1	0.83	1.07	1.68	2.44	3.75	3.51	2.89	2.80	2.56	2.08	1.45	1.00
Oklahoma 4	0.74	1.01	1.66	2.51	4.53	3.66	2.21	2.64	2.82	2.39	1.39	0.94
Oklahoma 7	0.98	1.19	1.80	2.63	4.81	3.68	2.34	2.55	3.05	2.82	1.51	1.20
Oklahoma 8	1.82	2.06	2.84	3.97	5.55	4.21	2.74	2.69	3.90	3.67	2.51	2.15
Oklahoma 9	3.07	3.22	4.06	5.00	6.22	4.25	3.76	3.40	4.24	3.98	3.86	3.77
Texas 1	0.45	0.58	0.87	1.39	2.80	2.71	2.54	2.44	2.16	1.63	0.78	0.58
Texas 2	0.80	1.03	1.23	2.16	3.66	3.04	2.10	2.29	2.87	2.40	1.26	0.98
Texas 3	1.89	2.22	2.48	3.71	4.79	3.30	2.26	2.17	3.14	3.26	2.51	2.35
Texas 4	3.74	3.50	3.77	4.57	4.99	3.86	3.37	2.82	3.36	3.65	4.17	4.58
Texas 5	0.47	0.46	0.36	0.61	1.14	1.48	1.97	1.93	2.06	1.36	0.55	0.54
Texas 6	1.09	1.45	1.40	2.39	3.50	2.66	2.01	2.17	3.15	2.74	1.58	1.31
Texas 7	2.01	2.19	2.06	3.03	4.12	3.33	2.29	2.49	3.84	3.46	2.51	2.39
Texas 8	3.47	2.98	2.97	3.36	4.60	4.43	4.57	3.88	4.69	4.01	3.83	4.02
Texas 9	1.11	1.21	1.18	2.01	3.19	2.68	1.79	2.17	3.42	2.42	1.37	1.20
Texas 10	1.30	1.20	0.97	1.64	2.83	2.58	1.73	2.12	4.56	2.66	1.34	1.25
Utah 7	0.75	0.69	0.83	0.72	0.70	0.47	0.91	1.16	1.02	1.15	0.68	0.71

For the most part, the spatial distributions of monthly averages closely resemble those for annual average precipitation. In Arizona, New Mexico, and Utah, winter precipitation is caused mainly by frontal activity associated with the general movement of Pacific Ocean storms across the country from west to east. As these storms move inland, most of the moisture is precipitated over mountain ranges, leaving areas east of the Continental Divide dry during this season. Winter (Fig. 8.2a) is the driest season in New Mexico except for the portion west of the Continental Divide. Summer rains in New Mexico are almost entirely the result of intense thunderstorms with the combination of Gulf moisture, orographic lifting, and surface heating.

Nearly half of the annual precipitation on average falls in the months of July and August (Fig. 8.2c) in New Mexico. Overall, Arizona, Utah, and New Mexico exhibit irregular patterns of seasonal precipitation complicated by the topography of the region. On smaller spatial scales, the precipitation gradients are very intense but this is not reflected in annual averages because of the overall dryness in this region. However, the seasonal difference is much larger in this area than in Texas and is based on the Continental Divide in winter seasons and the availability of Gulf and monsoonal moisture in the summer.

The seasonal precipitation gradients across Texas, Oklahoma, Arkansas, and Louisiana are stronger and more uniform on larger scales. In seasons other than summer, a general precipitation gradient exists with precipitation averages decreasing from East to West. The strongest gradient exists in January (Fig. 8.2a) with Pacific storm systems precipitating most of their moisture west of the Continental Divide before tapping into Gulf moisture in East Texas. The gradient relaxes a bit in the spring (Fig. 8.2b) with more available moisture in West Texas and the presence of a large-scale dryline. Air descending the eastern slopes of the Rockies warms and dries out as it sinks; creating a hot, dry, cloud-free zone that gives birth to the dryline.

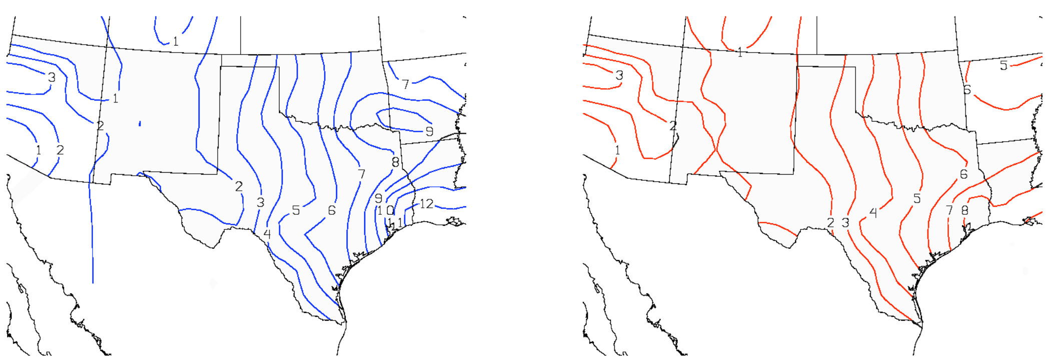
The gradient of summer precipitation (Fig. 8.2c) is the most complicated across Texas as much of the precipitation is due to surface heating rather than large-scale frontal features. More than one factor is responsible for the very tight precipitation gradient in Southeast Texas and Louisiana. One factor is the almost daily seabreeze shower activity that forms along the Gulf Coast and stays within roughly 100 miles of the coast, the area where the gradient slackens significantly on Figure 8.2c. Perhaps the most important contributors to the strong summer average precipitation gradient are the hurricanes and tropical systems that frequently batter the Gulf Coast. A very large percentage of the precipitation is cause by these systems.

Across the rest of Texas, summer precipitation is characterized by the interaction of Gulf moisture with surface heating combining to produce showers and thunderstorms. Upper level winds during the summer in Texas are generally light, so these precipitation events are generally isolated and can produce flooding rainfalls in one area while an area in close proximity can remain completely dry. Overall, a minimum in our domain average July precipitation (Fig. 8.2c) exists in West Central Texas. This area receives less moisture from the Gulf of Mexico than areas to the east and less monsoonal moisture than areas further west.

The weakest gradient in precipitation across Texas is that in autumn (Fig. 8.2d). This is a transition period from the mainly convective storms across most of Texas with seabreeze influences in Southeast Texas to a time with more frontal influences as seen in the winter. An interesting feature of Figure 8.2d is the appearance of discontinuities in the precipitation gradient across North Texas near the Red River. Frequently, October frontal boundaries stall out in North Texas and Southern Oklahoma. This is typified by the small bullseye of 4" precipitation in Southeastern Oklahoma and Southwestern Arkansas.

c. USHCN Dataset Precipitation Variances

Differences in the precipitation variance values between the actual and interpolated time sereis datasets are important to adjust for in creating accurate long-term records at each COOP station. Analysis of these variances for both the actual and interpolated USHCN station time series datasets appear in Figure 8.3. An intracomparison within each map demonstrates the spatial differences throughout Texas, New Mexico, and surrounding states. Comparisons between the two maps demonstrate the differences in variance magnitudes between the actual and interpolated USHCN time series datasets for each station. Figure 8.3 contours the variances for the actual (a) and interpolated (b) USHCN monthly variances.



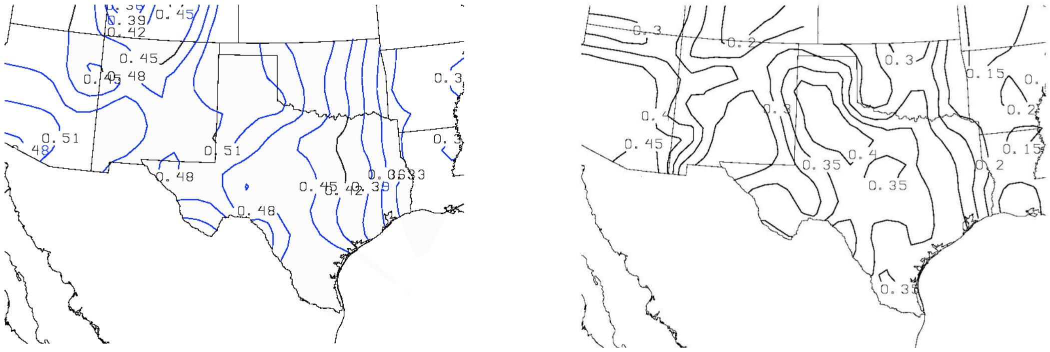
(a) UAM **(b) UIM**

FIG 8.3. Maps of USHCN variances using the UAM dataset (a) and the UIM dataset (b).

The spatial distributions of monthly precipitation variance are similar when compaing the actual and interpolated USHCN datasets. Generally, the variance decreases from East to West, with a minimum in monthly variances in New Mexico and the largest monthly variances in Louisiana. The main difference between the two datasets is the magnitudes of the variances in the actual dataset are larger than variances in the interpolated dataset.

The differences in variance are the driving force behind the creation of a third dataset that acts to replicate the variance of the actual USHCN monthly precipitation variances but has the data availability of the interpolated dataset. A spatial distribution map of variances for the third variance-adjusted dataset for both USHCN and COOP stations would be nearly identical to that in Figure 8.3 (a). This is because the variance at each station is an average of the variances from the nearest two USHCN stations.

Another measure of variability would be the normalized variance which takes the variance of a distribution divided by its mean. Figure 8.4 is a spatial plot of the UAM dataset normalized variance (a) and the UIM normalized variances. The normalized variance takes the variance of a particular time series and divides by the sum of the squared values in the time series (Wilks 2006).



(a) UAM **(b) UIM**

FIG 8.4. Maps of USHCN normalized variances using the UAM dataset (a) and the UIM dataset (b).

The results from Figure 8.4 show the variance of precipitation normalized by how much precipitation an area actually receives. This statistic has important implications for drought because higher values of normalized variance would mean an area is more likely to be at either end of its precipitation distribution. From both maps on Figure 8.4, it is plain to see the largest magnitudes of normalized variance are in West Texas and New Mexico.

9. PRECIPITATION TREND ANALYSIS

a. Introduction

The numerous datasets in this project (Table 8.1) allow for intercomparisons of long-term precipitation trends. The first USHCN dataset of interest uses actual long-term USHCN station data, the ULY dataset, and actual data from all USHCN stations, the UAY dataset. The second USHCN dataset includes interpolated data from all 221 stations in this study, or the UIY dataset.

Also of interest are the precipitation trends of the actual COOP precipitation data, the CAY dataset, and the interpolated COOP data, the CIY dataset. Following these analyses will be the precipitation trends of the CQY and CqY datasets. The high-quality USHCN stations used in the interpolation process of these two datasets are those located in New Mexico and Texas and the designation of these stations as high-quality is included in Appendix C. The CQY dataset looks at the interpolated data using the high-quality stations while the CqY analyzes the variance-adjusted time series at stations using only the high-quality interpolated data.

Dot plot maps are included in the analysis of the USHCN data with the dots corresponding to the magnitude of the precipitation trend. Positive trends are denoted by blue dots and red dots denote negative precipitation trends. Time series analyses using running 12-month precipitation totals are used to display the variation in trends across the 20th century for the four COOP datasets.

The trend for a period of time is determined by fitting a least squares regression line to the time series data for a particular USHCN station time series or COOP climate division averaged time series. Since these analyses are focused on the entirety of the 20th century, the units will be denoted as inches of precipitation per century.

b. Dot Plots of 20th Century Precipitation Trends Using USHCN Datasets

1) USHCN long-term stations (ULY)

The first dot plot analysis includes USHCN stations with data dating back to the start of the 20th century. Ninety-six of the 221 USHCN stations fit this criterion and are labeled by boldface type in Appendix A. Figure 9.1 contains a dot plot map of the ULY dataset station precipitation trends for the 20th century.

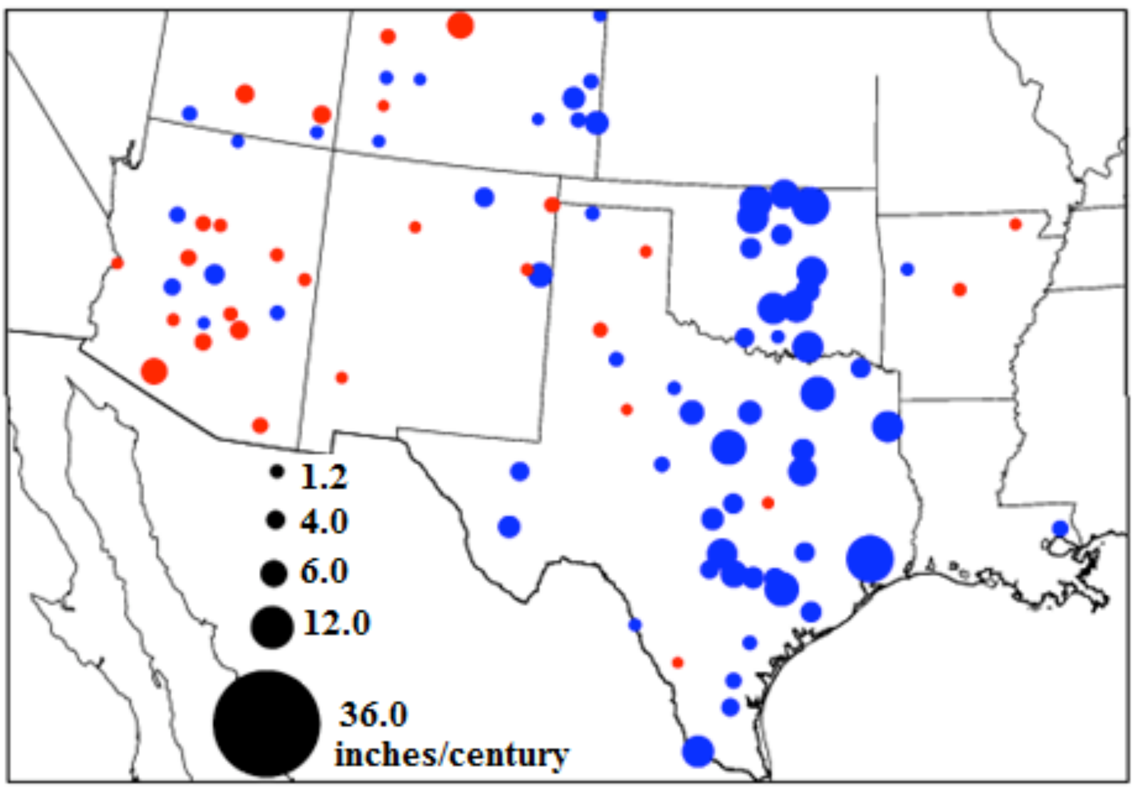


FIG 9.1. Dot plot map of precipitation trends over the 20th century using the ULY dataset. The legend refers to the trend in inches per century.

2) USHCN actual data (UAY)

Several of the USHCN stations have records beginning in the year 1948, especially in Louisiana and New Mexico. For this reason, this dot plot analysis (Fig. 9.2) examines the time period from 1948-2000 for the actual USHCN data, the UAY dataset, to include the remainder of the stations and to show the differences between the entire 20th century precipitation trends and the trend for the latter half of the century.

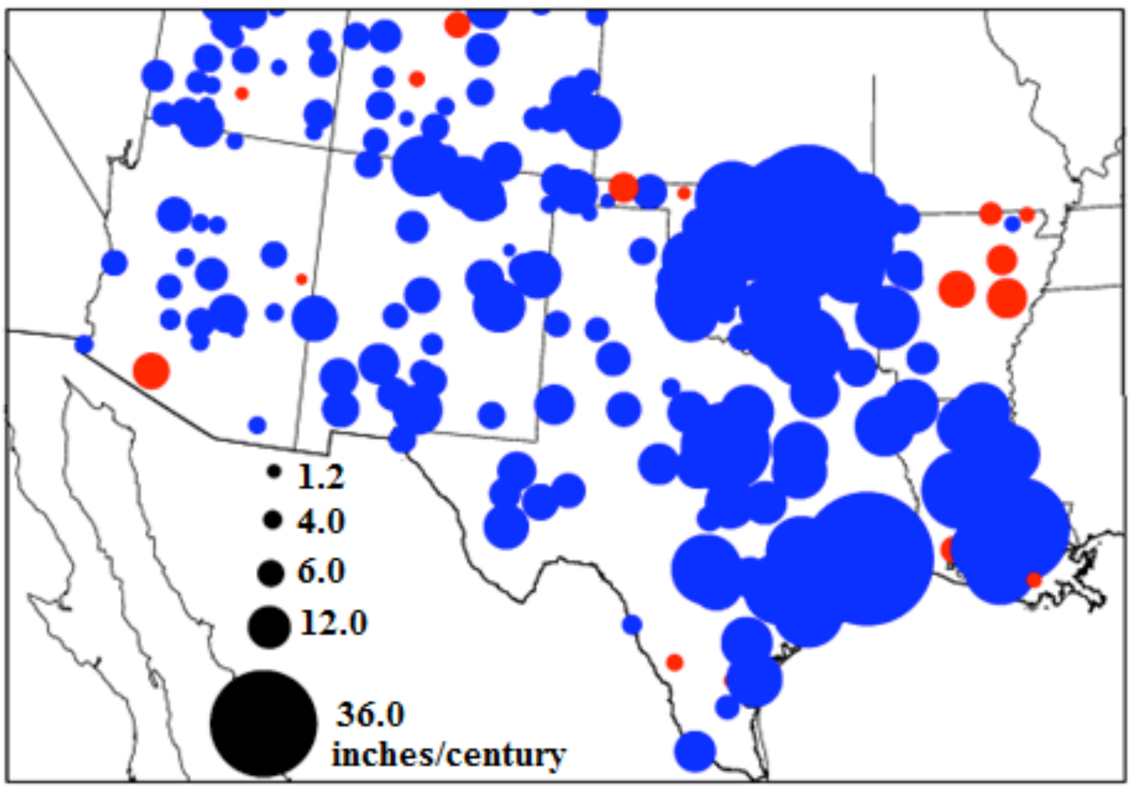


FIG 9.2. Dot plot map of precipitation trends from 1948-2000 using the UAY dataset. The legend refers to the trend in inches per century.

3) Interpolated data from all USHCN stations (UIY)

This set of analyses contains those from all 221 USHCN stations in our study and the interpolated data from each of these stations. The interpolated data are used in these analyses since several of the USHCN stations have shorter than century-long actual climate records. The 20th century dot plot map of the UIY dataset station precipitation trends are shown in Figure 9.3.

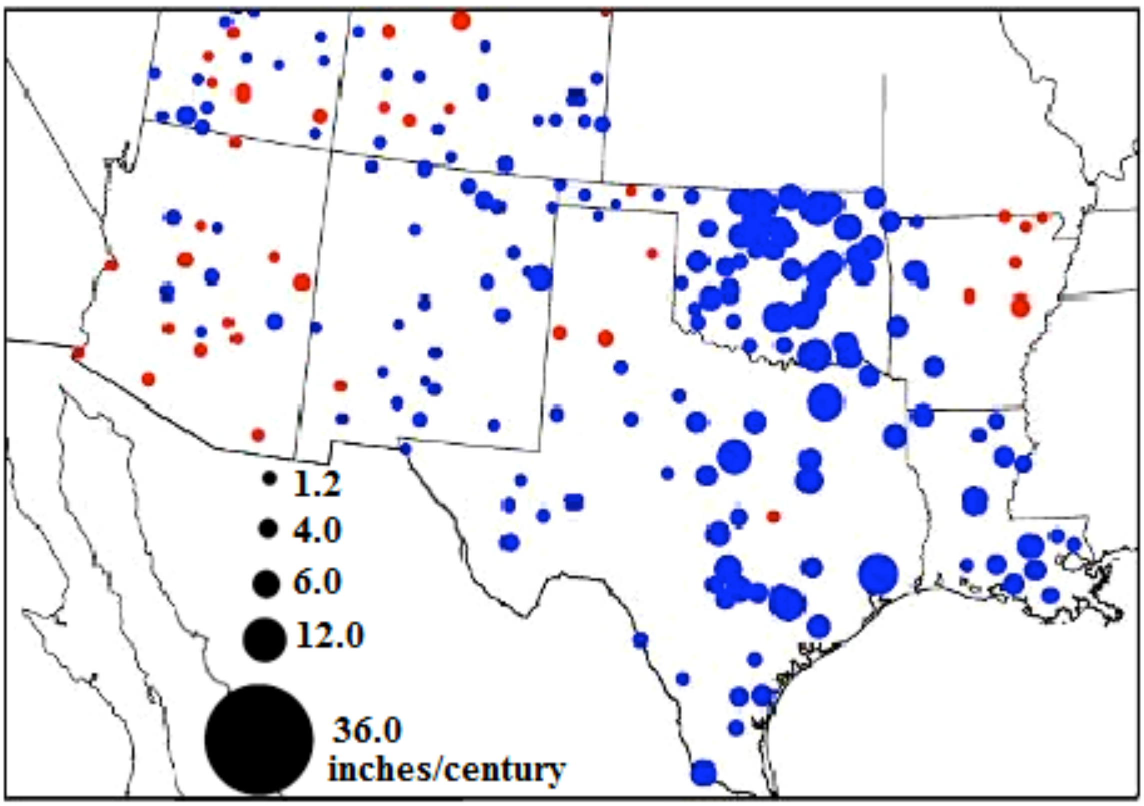


FIG 9.3. Dot plot map of precipitation trends over the 20th century using the UIY dataset. The legend refers to the trend in inches per century.

4) Analysis of USHCN dot plots

The same stations in each of the three dot plot maps, Figures 9.1-9.3, show more increasing than decreasing trends. However, the major difference is in the magnitudes of values in Figure 9.2 compared to the other two plots. There are two possible causes for the differences. The first is an actual change in precipitation trends that shows a greater magnitude in positive trend (Fig. 9.2) over most of the domain during the latter half of the 20th century.

The second possibility could be the interpolation process in Fig. 9.3 tempers the trend because of its disposition to reduce extremely high values of precipitation. However, the similarity between the collocated stations in Figures 9.1 and 9.3, the two dot plots showing trends over the entire 20th century, suggest the difference is due to actual changes in the precipitation trends. There was a widespread period of drought near the beginning of this period (1948-2000) in the 1950s and wet conditions over most of Texas and New Mexico at the end of the period.

c. Time Series Analyses of COOP Datasets

Because of the spatial density of the COOP network it would be difficult to replicate the dot plot studies done for the three USHCN datasets. This section contains analyses of the four COOP datasets in the form of time series graphs showing 12-month precipitation anomalies.

The anomalies were originally calculated on a station-by-station basis, before being averaged across each climate division for every month. All the available data for each month in a given climate division were utilized. For each dataset, the analyses contain precipitation trends over the entirety the 20th century, broken down into the regions denoted by Figure 7.3.

Each climate division time series includes the raw data of running 12-month anomalies. The raw data are useful for showing the extremes in each climate division's distribution and the general trends in precipitation across the 20th century. In each graph, the right-side vertical axis refers to the running 12-month averages over the entire region, either the East Texas region, the West Texas region, or the New Mexico region. The colored lines refer to the running 12-month climate division averages with the climate division label on the left-side vertical axis. The thick red line in each graph is the climate region smoothed 5-year centered-average of the 12-month running precipitation totals.

1) COOP climate-division averages – interpolated data (CIY)

These analyses are for the CIY dataset, containing interpolated data for COOP stations that used all 221 USHCN stations in the interpolation process. In this section and the next section, three different time series graphs show the 12-month precipitation running totals for the 33 climate divisions. The running climate division averages are grouped according to the three regions display on Figure 7.3, the first group represents the East Texas region (Fig. 9.4), the second represents the West Texas region (Fig. 9.5), and the third represents the New Mexico region (Fig. 9.6).

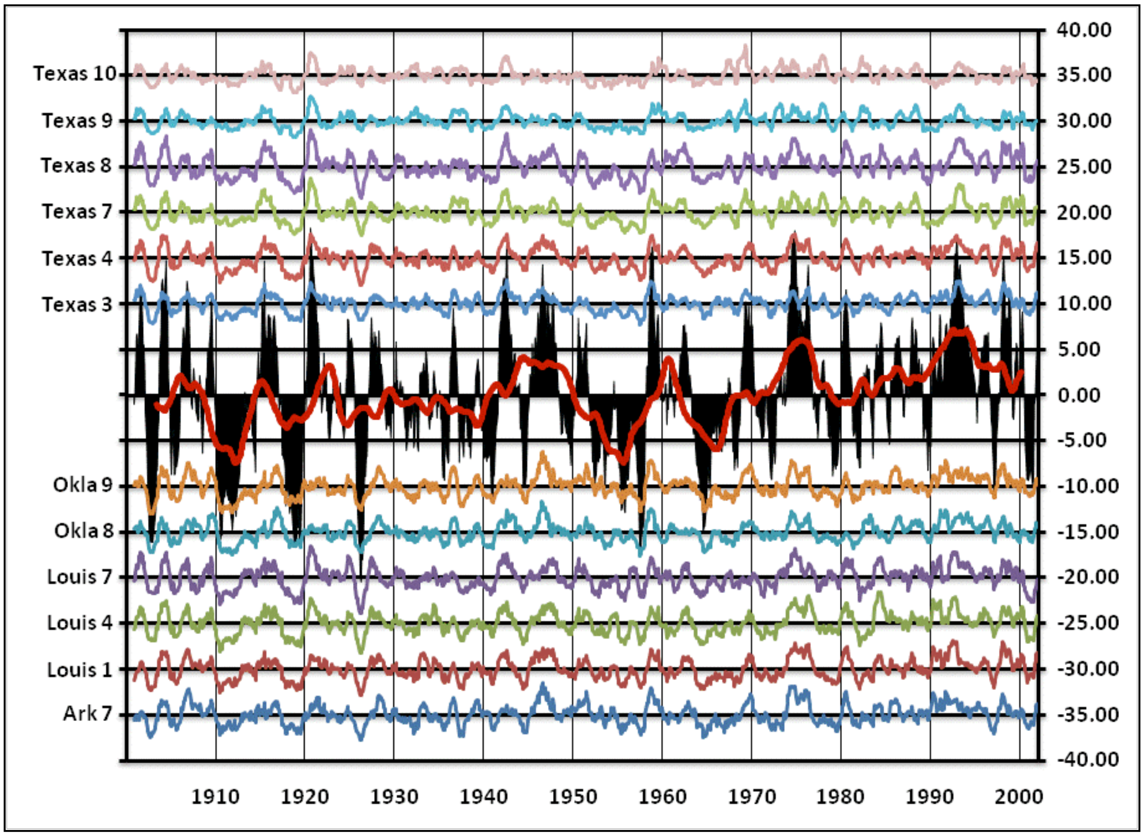


FIG 9.4. Time series climate division averages from the CIY dataset for all the climate divisions in the East Texas region on Figure 7.3, with a magnitude of 40" between horizontal lines for climate division data. The black indicates an anomaly for the climate region averages at a given time with the red line a 5-year centered-average.

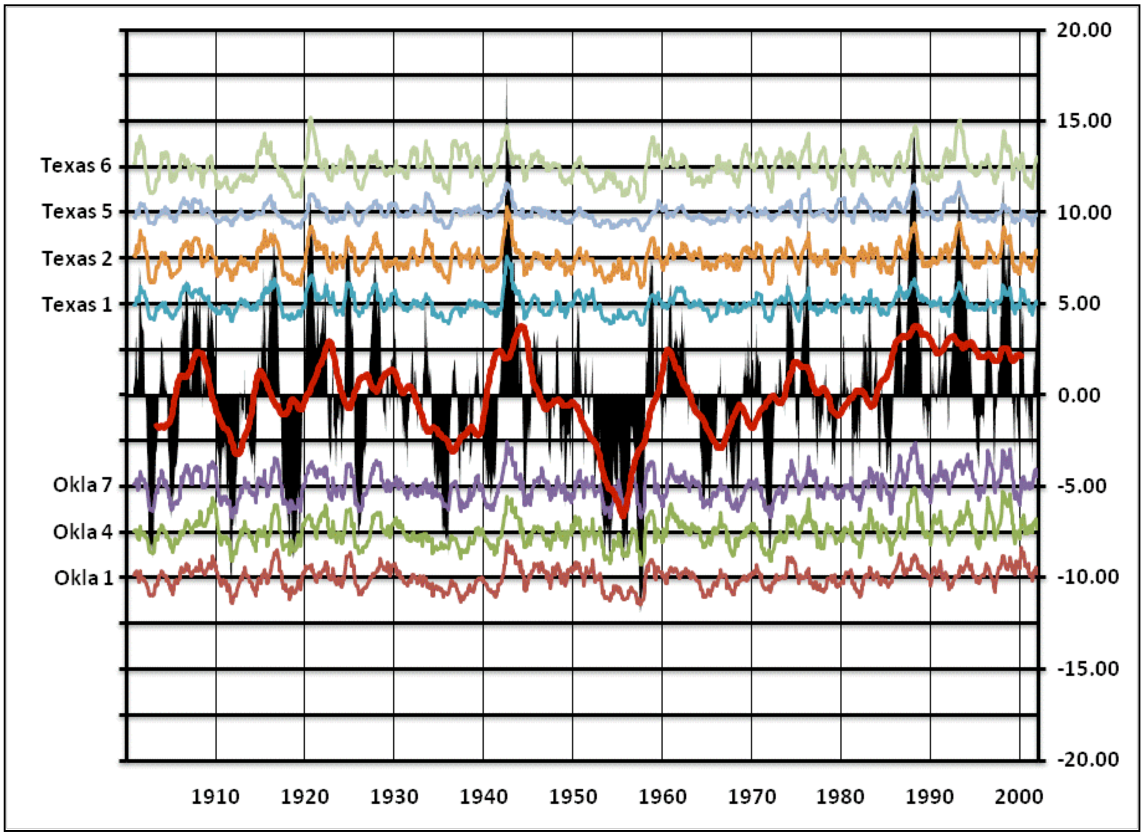


FIG 9.5. Time series climate division averages from the CIY dataset for all the climate divisions in the West Texas region on Figure 7.3, with a magnitude of 20” between horizontal lines for climate division data. The black indicates an anomaly for the climate region averages at a given time with the red line a 5-year centered-average.

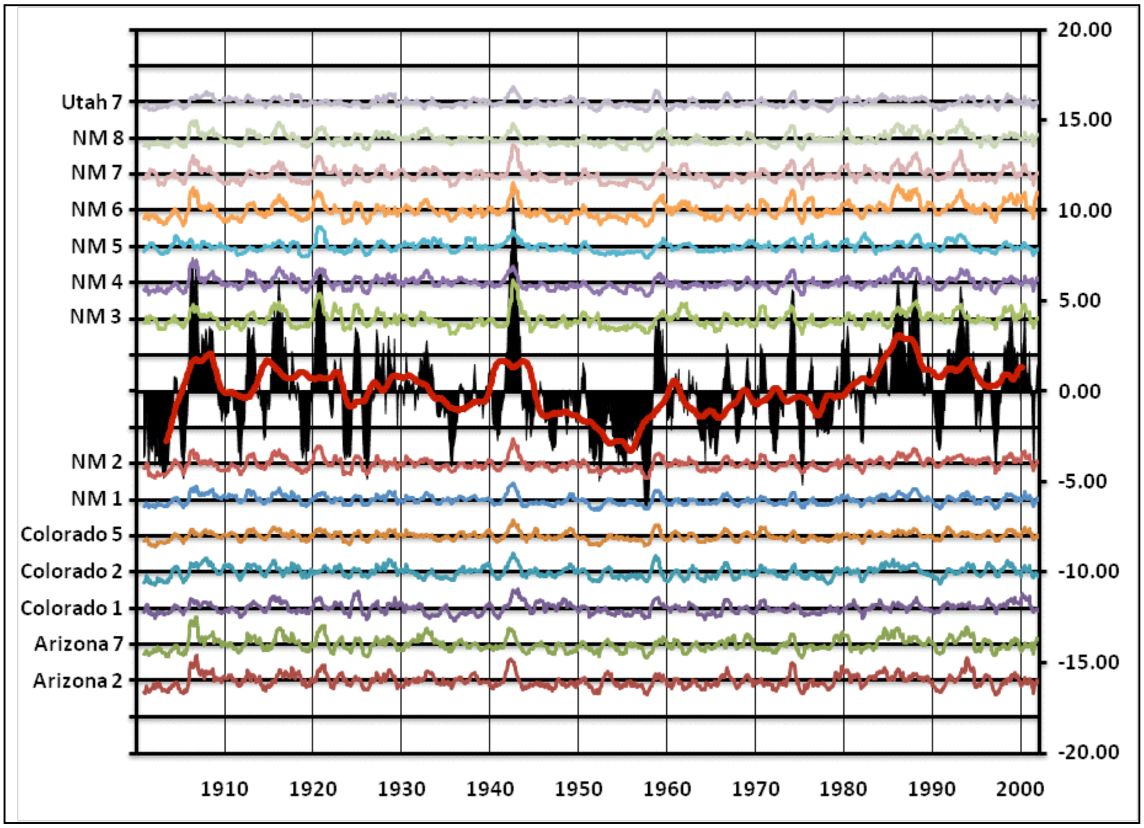


FIG 9.6. Time series climate division averages from the CIY dataset for all the climate divisions in the New Mexico region on Figure 7.3, with a magnitude of 20" between horizontal lines for climate division data. The black indicates an anomaly for the climate region averages at a given time with the thick red line indicative of a 5-year centered-average.

2) COOP climate division averages – actual data (CAY)

These analyses contain precipitation trends for the actual climate division COOP data. The time series graphs (Figs. 9.7-9.9) for these datasets are grouped according ot the three regions in Figure 7.3. This CAY dataset represents data actually recorded by the observers and equipment with the exception of observations eliminated in the quality control procedures by NCDC and the quality control procedures described in Section 4.

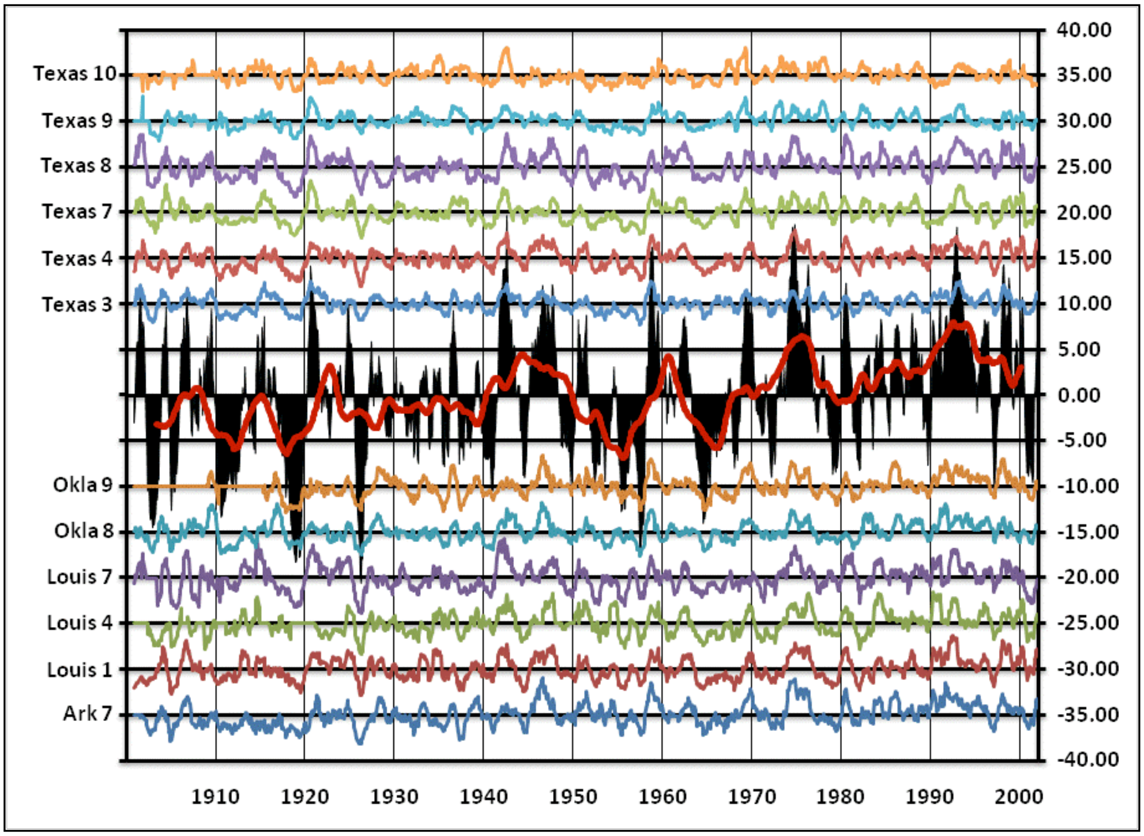


FIG 9.7. Time series climate division averages from the CAY dataset for all the climate divisions in the East Texas region on Figure 7.3, with a magnitude of 40" between horizontal lines for climate division data. The black indicates an anomaly for the climate region averages at a given time with the thick red line indicative of a 5-year centered-average.

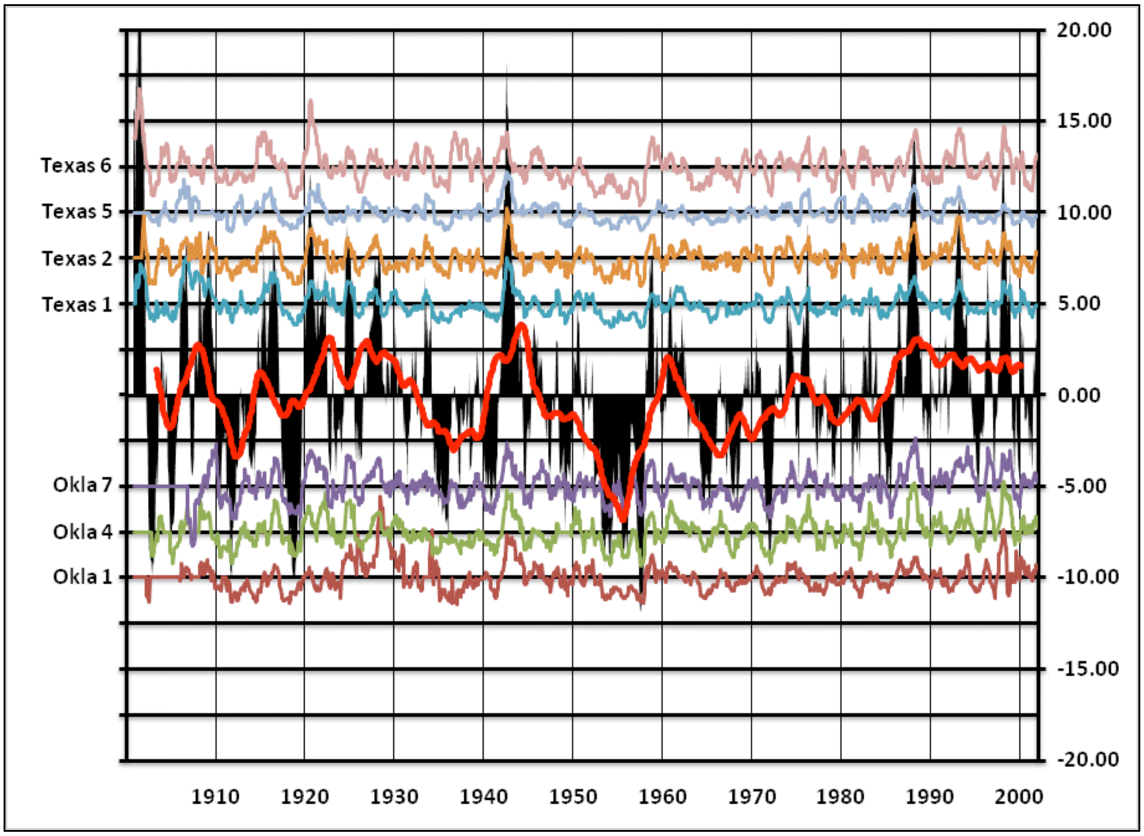


FIG 9.8. Time series climate division averages from the CAY dataset for all the climate divisions in the West Texas region on Figure 7.3, with a magnitude of 20” between horizontal lines for climate division data. The black indicates an anomaly for the climate region averages at a given time with the red line a 5-year centered-average.

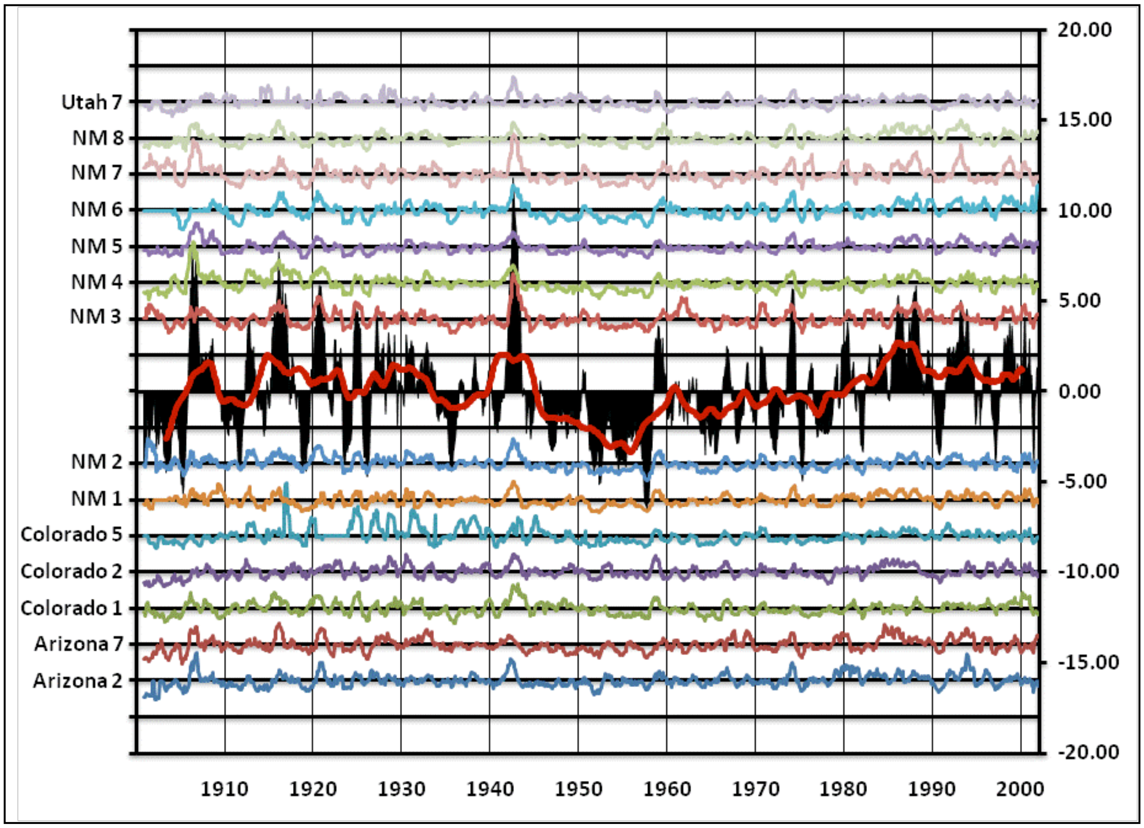


FIG 9.9. Time series climate division averages from the CAY dataset for all the climate divisions in the New Mexico region on Figure 7.3, with a magnitude of 20" between horizontal lines for climate division data. The black indicates an anomaly for the climate region averages at a given time with the thick red line indicative of a 5-year centered-average.

3) COOP climate division averages – high-quality interpolated data (CQY)

The following analyses contain COOP interpolated data using only high-quality stations, the CQY dataset, denoted by italics in Appendix A in the interpolation process. Since the metadata study was limited to stations only in New Mexico and Texas, the results in this section and the following section will be limited to climate divisions in New Mexico and Texas. The figures in this section will contain the time series anomaly plots of 12-month running precipitation totals for the East Texas (Fig. 9.10), West Texas (Fig. 9.11), and New Mexico (Fig. 9.12) climate divisions using the CQY dataset.

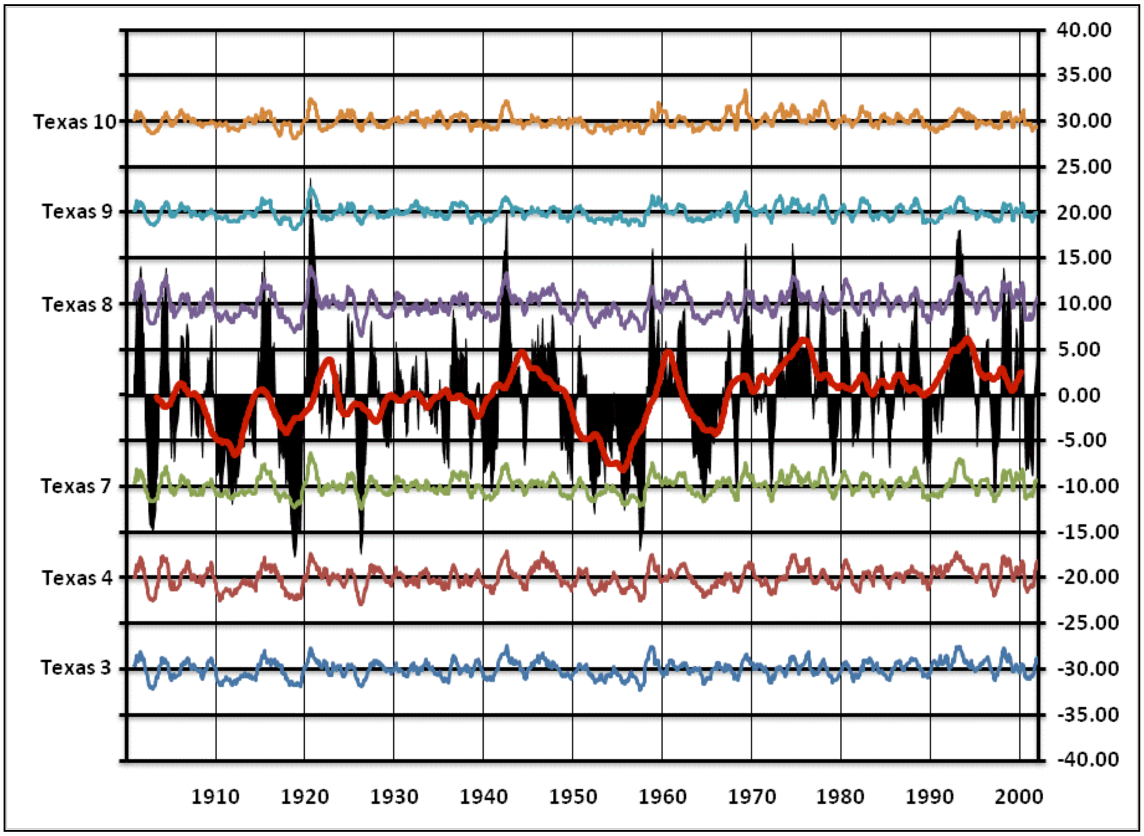


FIG 9.10. Time series climate division averages from the CQY dataset for all the climate divisions in the East Texas region on Figure 7.3, with a magnitude of 40" between horizontal lines for climate division data. The black indicates an anomaly for the climate region averages at a given time with the red line a 5-year centered-average.

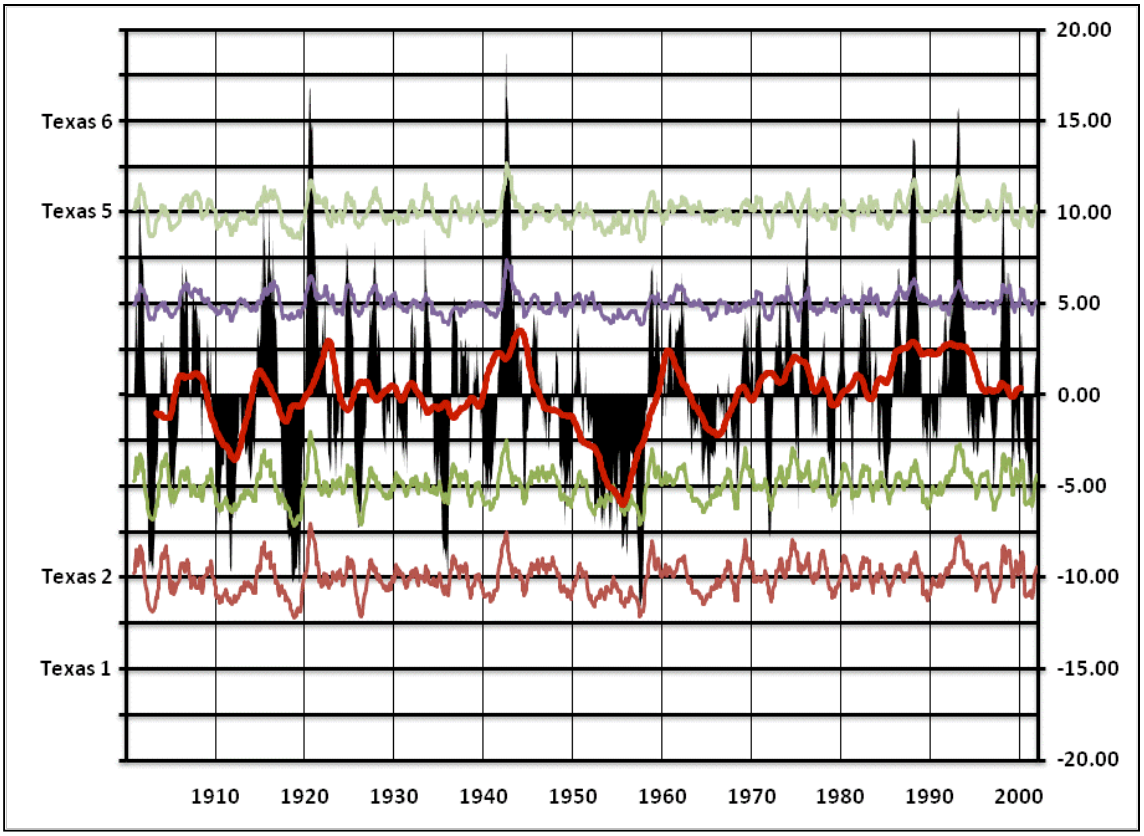


FIG 9.11. Time series climate division averages from the CQY dataset for all the climate divisions in the West Texas region on Figure 7.3, with a magnitude of 20" between horizontal lines for climate division data. The black indicates an anomaly for the climate region averages at a given time with the thick red line indicative of a 5-year centered-average.

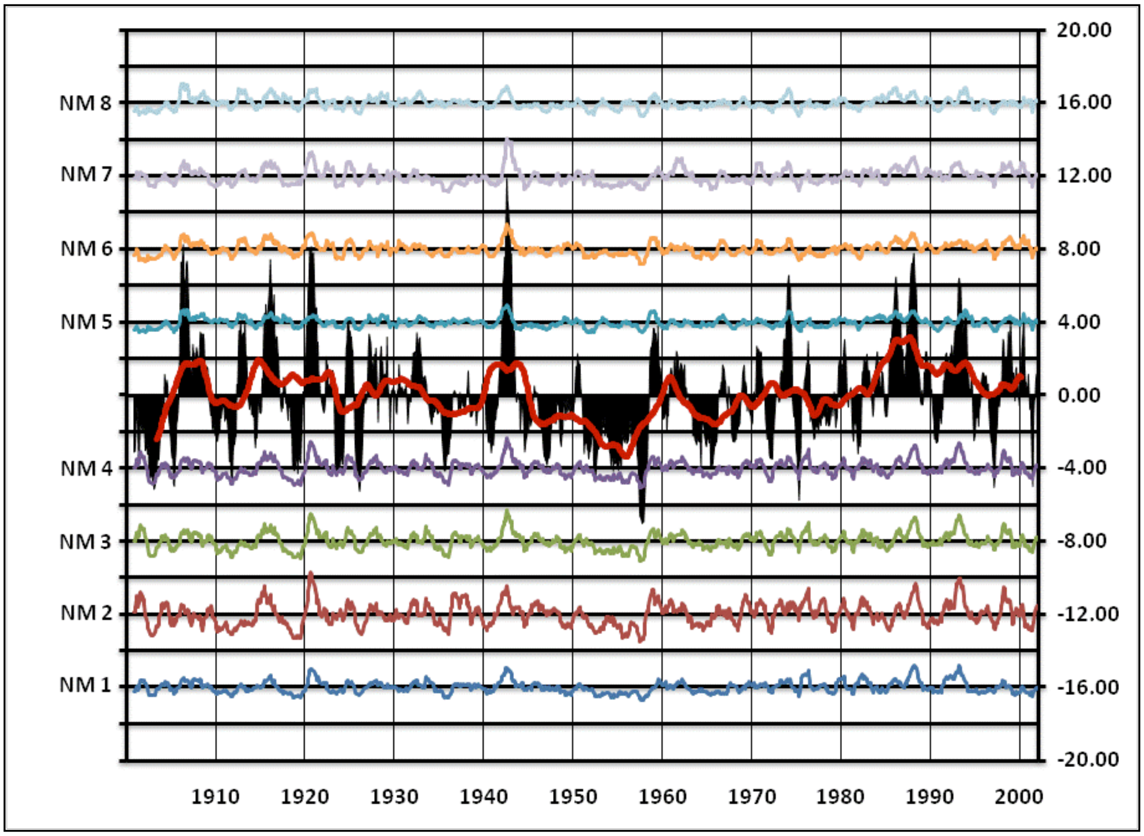


FIG 9.12. Time series climate division averages from the CQY dataset for all the climate divisions in the New Mexico region on Figure 7.3, with a magnitude of 20" between horizontal lines for climate division data. The black indicates an anomaly for the climate region averages at a given time with the thick red line indicative of a 5-year centered-average.

4) COOP climate division averages – high quality variance-adjusted data (CqY)

The following section the CqY dataset, containing variance-adjusted time series data from high-quality USCHN stations used in the interpolation process. Figure 9.13 contains 12-month precipitation anomalies for East Texas, anomalies for West Texas are in Figure 9.14, and Figure 9.15 has the anomalies for all eight New Mexico climate division averages. This dataset is important because it is temporally complete, has had the variance of its time series adjusted to those characteristic of long-term USHCN stations, and uses high-quality USHCN stations in the interpolation process.

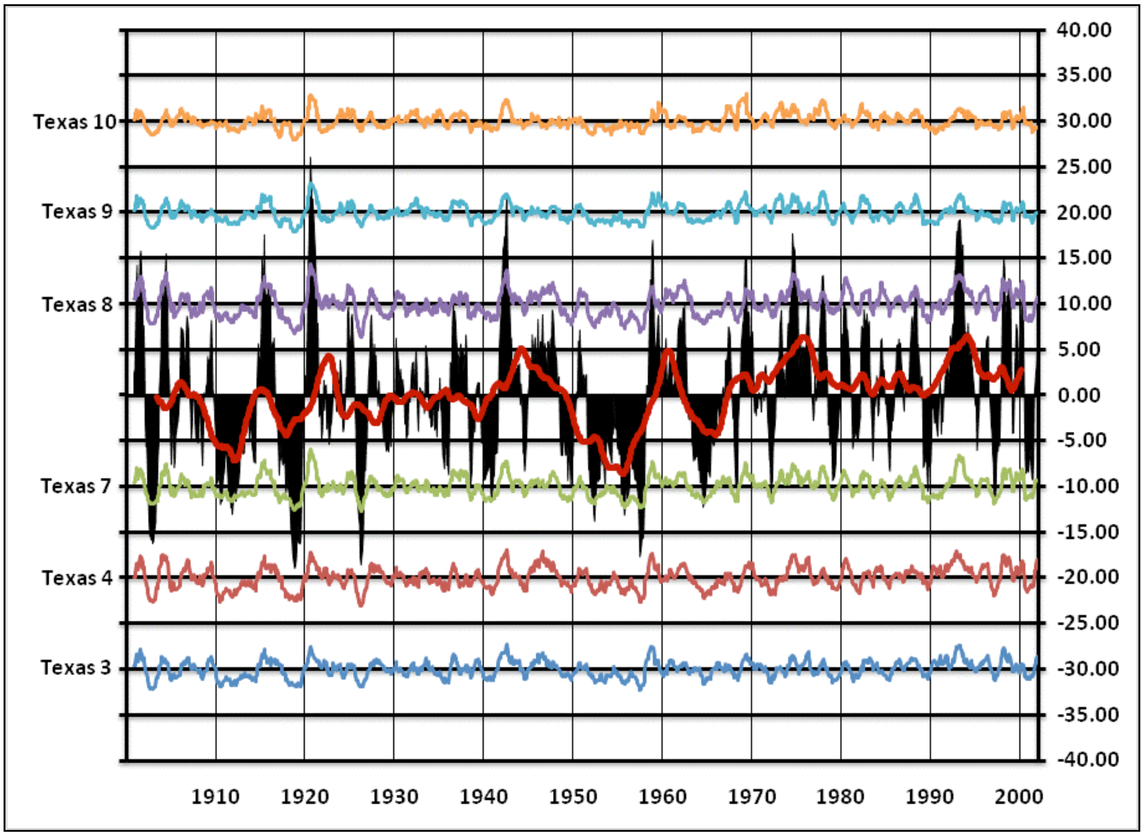


FIG 9.13. Time series climate division averages from the CqY dataset for all the climate divisions in the East Texas region on Figure 7.3, with a magnitude of 40" between horizontal lines for climate division data. The black indicates an anomaly for the climate region averages at a given time with the thick red line indicative of a 5-year centered-average.

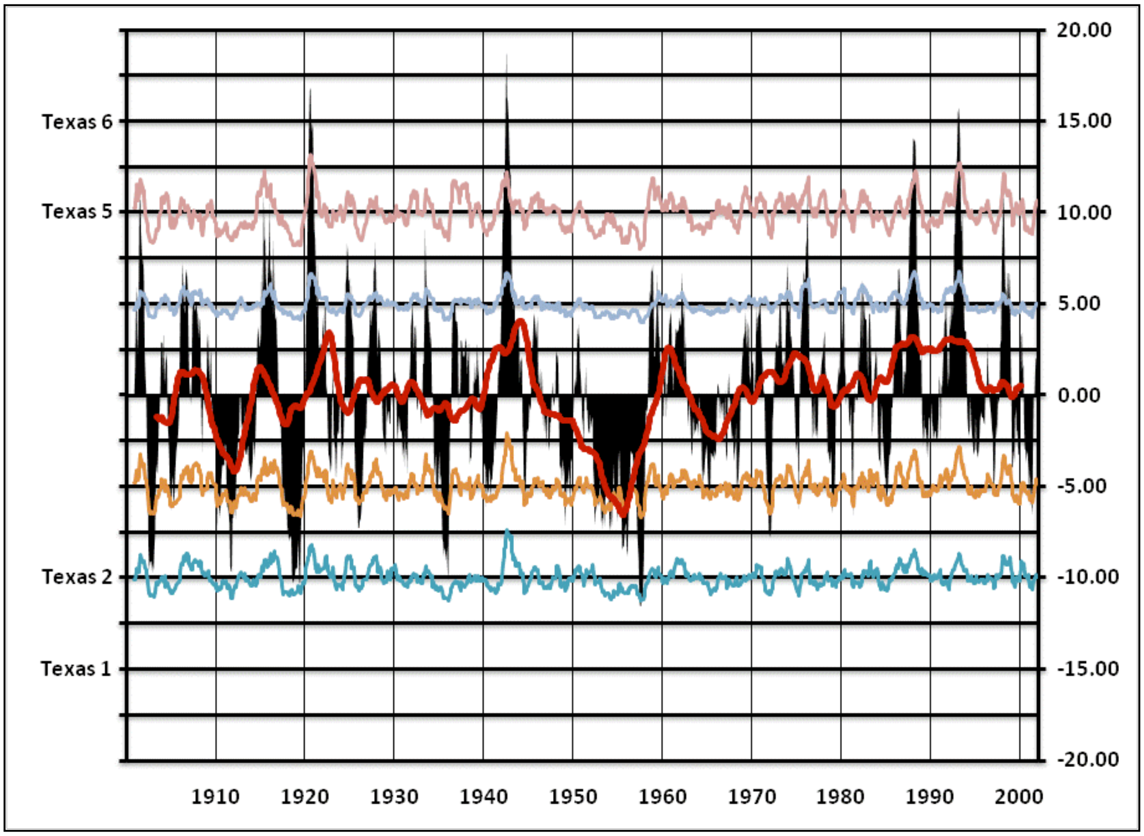


FIG 9.14. Time series climate division averages from the CqY dataset for all the climate divisions in the West Texas region on Figure 7.3, with a magnitude of 20" between horizontal lines for climate division data. The black indicates an anomaly for the climate region averages at a given time with the thick red line indicative of a 5-year centered-average.

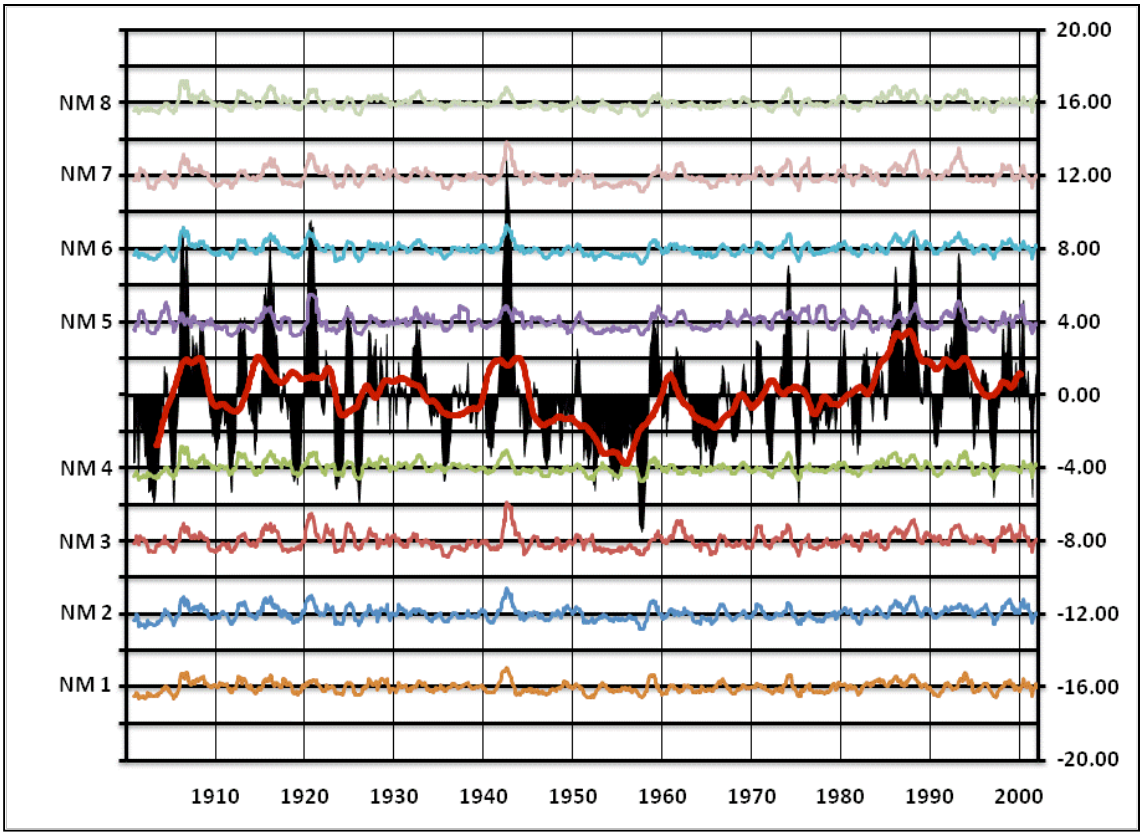


FIG 9.15. Time series climate division averages from the CqY dataset for all the climate divisions in the New Mexico region on Figure 7.3, with a magnitude of 20" between horizontal lines for climate division data. The black indicates an anomaly for the climate region averages at a given time with the thick red line indicative of a 5-year centered-average.

5) Analysis of COOP time series plots

A great deal of information can be derived from the 20th century time series analyses for the four COOP datasets. The most apparent characteristic shared by all twelve time series plots (Figs. 9.4-9.15) is the drought period in the 1950s. For the most part, there is a positive anomaly in precipitation toward the end of the 20th century. However, the New Mexico climate divisions and Texas climate divisions 9 and 10, both in the southern part of Texas, seem to show a decreasing trend in this time period. The decreasing trends in New Mexico are followed by relative maxima, a characteristic not in the time series for the two Texas climate divisions.

The time series plots suggest in all the figures that the driest stretches of 12-month precipitation occurred prior to the long-term drought of the 1950s. Following the drought of the 1950s through the end of the 20th century, there are few large magnitude and long-lasting dry periods. There is a signal in the three regions for a drought in the early 1960 which is strongest in the East Texas region. However, the longer-term trends in the individual climate divisions show this to be a small fluctuation in an overall increasing trend.

Most of the long-term individual climate division time series lines show a positive anomaly for most of the last quarter of the 20th century, with relative maxima in the late 1980s to early 1990s. However, the only climate division showing an increasing trend through the end of the 20th century was Oklahoma climate division 1, located in its panhandle. Focusing in on the CQY and CqY datasets for New Mexico and Texas, there appears to be a lag in the relative maxima toward the end of the 20th century. The maxima in the New Mexico climate divisions occur in the late 1980s and in the early 1990s in the East Texas and West Texas climate divisions.

Taking a closer look at the 1910s shows the precipitation anomalies in the two Texas regions to show a major peak surrounded by two major relative minima. This anomaly pattern would show up as a minimum in the long-term signal for each climate division at the same time a peak is occurring in the 12-month data. This is a prime example of opposite signals being detected simultaneously on different time scales.

Mentioned earlier was the relative lack of drought strength and longevity in the latter half of the 20th century compared to the first half. Visual analysis of the time series plots shows the frequency of pluvial conditions in each of the three climate regions denoted by Figure 7.3 to be increasing. However, there does not appear to be any discernible differences in the magnitudes of these periods when compared to the pluvial periods in the early 20th century.

The magnitude of the smoothed curve for most of the climate divisions has a maximum in the late 20th century. The combination of these two observations would suggest an increase in the frequency of 12-month pluvial periods rather than an increase in the severity of these periods. Further analysis on drought and pluvial conditions will be conducted in Section 10.

d. Precipitation Trends in the 20th Century Using COOP Datasets

Because of the large number of COOP stations, the century-long precipitation trends for the COOP datasets will be displayed as climate division averaged trends in a tabular format following the time series analyses. The Table 9.1 will show the results for the CAY, CIY, CQY, and CqY climate division averaged 20th century trends. Also included on Table 9.1 are the percentage changes in the expected rainfall total in 2001 compared to the beginning of the time series in 1900.

TABLE 9.1. The 20th century precipitation trends for the 33 climate divisions in this study with the units as *inches of precipitation per century*. Also included is the percentage change in expected precipitation from 1901 to 2000.

	1900-2000 Precipitation Trend				**1900-2000 Percentage Change**			
	CIY	**CAY**	**CQY**	**CqY**	**CIY**	**CAY**	**CQY**	**CqY**
Arizona 2	-0.15	2.04	-	-	-0.98	14.49	-	-
Arizona 7	-0.05	2.30	-	-	-0.31	16.16	-	-
Arkansas 7	4.85	12.20	-	-	9.26	24.11	-	-
Colorado 1	0.66	0.01	-	-	4.34	0.05	-	-
Colorado 2	-0.22	1.11	-	-	-1.23	6.72	-	-
Colorado 5	0.45	-2.39	-	-	3.86	-21.51	-	-
Louisiana 1	6.02	6.01	-	-	12.09	12.17	-	-
Louisiana 4	6.12	10.51	-	-	11.38	19.89	-	-
Louisiana 7	4.61	6.78	-	-	7.83	11.45	-	-
New Mexico 1	-0.06	0.40	0.48	1.67	-0.56	3.66	4.33	15.08
New Mexico 2	2.16	-0.51	0.93	0.88	13.32	-3.09	5.75	5.43
New Mexico 3	1.00	0.77	1.36	1.78	6.25	4.88	8.54	11.17
New Mexico 4	-0.02	-1.34	-0.24	-0.43	-0.11	-9.26	-1.71	-3.03
New Mexico 5	0.58	2.04	0.91	0.72	5.76	21.59	9.11	7.61
New Mexico 6	0.42	2.00	1.12	1.33	2.39	12.32	6.34	7.60
New Mexico 7	1.05	-0.40	0.87	0.99	7.68	-2.91	6.36	7.28
New Mexico 8	0.94	2.69	0.59	0.84	8.03	24.46	5.08	7.23
Oklahoma 1	2.00	1.70	-	-	7.80	7.43	-	-
Oklahoma 4	3.08	2.17	-	-	11.67	8.08	-	-
Oklahoma 7	3.16	4.77	-	-	11.12	16.89	-	-
Oklahoma 8	4.91	5.59	-	-	12.91	15.07	-	-
Oklahoma 9	6.94	9.92	-	-	14.28	20.93	-	-
Texas 1	0.03	-2.08	0.39	0.26	0.16	-10.64	2.05	1.35
Texas 2	1.72	0.60	1.58	1.62	7.28	2.57	6.68	6.87
Texas 3	4.32	3.63	4.51	4.84	12.76	10.78	13.33	14.29
Texas 4	5.66	7.13	5.94	6.08	12.31	15.75	12.90	13.20
Texas 5	0.57	-0.84	0.46	0.32	4.56	-6.55	3.64	2.54
Texas 6	2.85	-0.38	2.99	3.22	11.34	-1.46	11.90	12.82
Texas 7	4.35	5.33	4.38	4.79	12.97	15.88	13.04	14.27
Texas 8	6.51	9.80	6.22	5.99	14.00	21.32	13.38	12.88
Texas 9	1.74	2.43	1.90	2.32	7.37	10.47	8.05	9.92
Texas 10	2.72	2.51	2.73	3.23	11.39	10.36	11.43	13.57
Utah 7	-0.10	-0.66	-	-	-1.08	-6.87	-	-

The overriding theme of Figure 9.16 is an increase in mean precipitation for most of the climate divisions, regardless of the dataset. However, there is some disagreement for particular climate divisions between the different datasets, particularly in the western part of our domain. An example of this disagreement would be in New Mexico climate division 2, which had a 13.32% increase in precipitation for the CIY dataset and a 3.09% decrease in the CAY dataset. Another is in Texas climate division 5, which had a 4.56% increase in precipitation for the CIY dataset and a 6.55% decrease in the CAY dataset.

Focusing on Texas climate division 5, Figure 9.8 shows the precipitation anomalies in the CAY dataset for the West Texas region. A significant positive anomaly occurs at the beginning of the 20th century in the CAY dataset not seen in the CIY dataset. This disagreement is likely caused by the discrepancy in the lack of available data at the beginning of the 20th century in the CAY dataset. Thought this may have been a pluvial period, the CAY precipitation anomaly for Texas climate division 5 was artificially enhanced during this period as only climatologically wet stations were available for the climate division average in the CAY dataset.

However, there is some question as to the precision of the interpolated values toward the beginning of the 20th century. There is a lack of long-term USHCN stations in West Texas and New Mexico, lading to a lack of data toward the beginning of the 20th century. The neighboring stations used in the interpolation process during this time would be a much greater distance from target stations than later in the 20th century. Therefore, the interpolated values at the beginning of the 20th century in West Texas and New Mexico may not be entirely reflective of precipitation patterns occurring in these areas.

10. PRECIPITATION DISTRIBUTION EXTREMES AND DROUGHT

a. Introduction

The extremes and long-term trends in the 12-month precipitation distributions will form the cornerstone of our drought analyses. Gibbs and Maher (1967) grouped monthly precipitation occurrences into deciles so that, by definition, “much lower than normal” weather cannot occur more often than 20% of the time. The classification of drought using this approach provides an accurate statistical measurement of precipitation given long climatic data records that are included in this study.

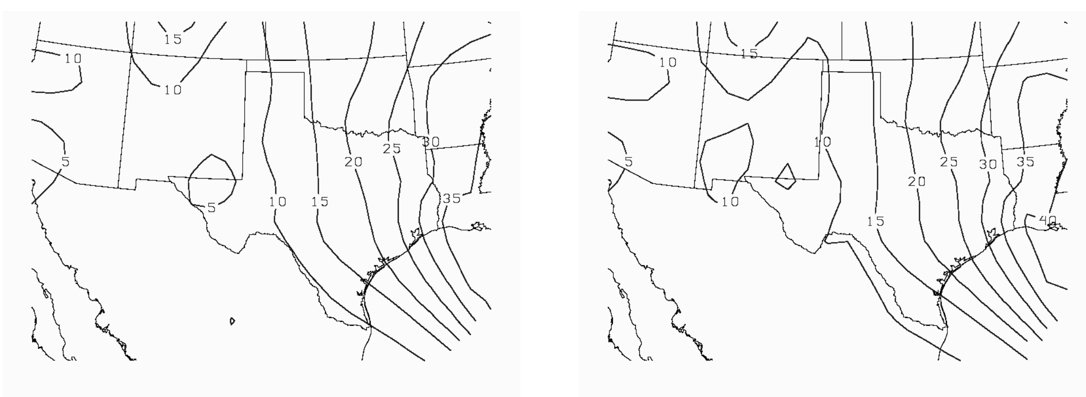
Also, the analyses on drought will focus on the datasets thought to provide the most accurate and long-term representation of precipitation trends. The drought analyses will focus on the long-term USHCN stations, the ULM and ULY datasets. The variance-adjusted values interpolated only using USHCN station deemed homogeneous in Appendix A will be used for the COOP stations; the CqM and the CqY datasets.

As for the precipitation distributions, the shape parameter remains relatively stable, but the scale parameter is variable spatially and temporally (Groisman et al., 1999). When the time period gets longer, the gamma distribution representing the precipitation distribution at a particular station approaches a normal distribution. Precipitation analyses will largely focus on the tail-ends of the gamma precipitation distributions and their effects on drought statistics. Karl and Knight (1998) found that across the United States, the total proportion of annual precipitation derived from extreme events has increased relative to moderate precipitation.

This implies that a shift in the mean precipitation will have the most influence on extreme precipitation totals. For each station, it is assumed the mean monthly precipitation is equal for the three different time series based on the interpolation process for the second dataset and gamma distribution fitting for the third dataset. However, it is the size and shape parameters of the gamma distribution that cause the three time series data variances to be unequal.

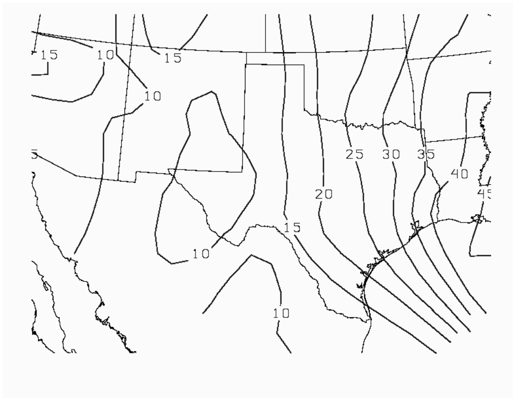
b. Precipitation Averages for Gamma Distribution Using Extreme CDF Values

Using the USHCN data, the gamma distributions were calculated for each of the three different time series at each station. One can determine precipitation totals equivalent to certain CDF values of precipitation and the spatial properties among the different types of precipitation time series. The following maps (Fig. 10.1) contain the USHCN precipitation totals from the ULY dataset representing CDF values of 0.02, 0.05, 0.10, 0.20, 0.80, 0.90, 0.95, and 0.98 for 12-month precipitation totals.

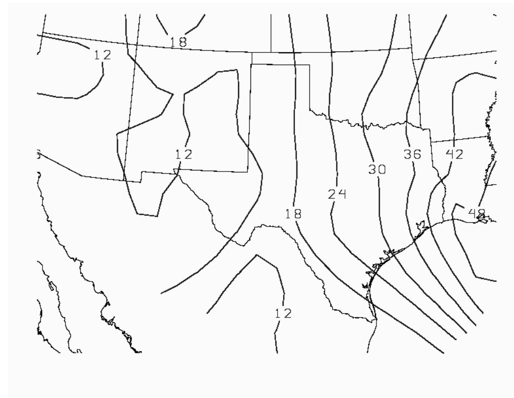


(a) 0.02 **(b) 0.05**

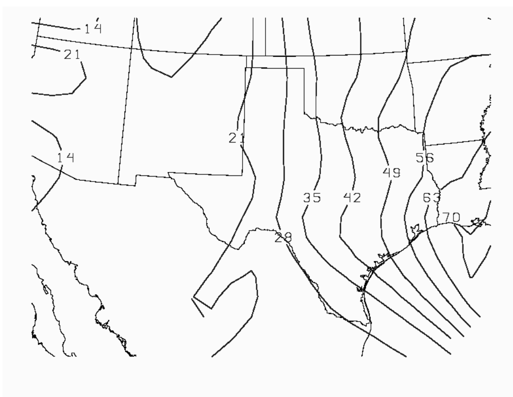
FIG 10.1. Maps of 12-month precipitation CDF values for 0.02 (a), 0.05 (b), 0.10 (c), 0.20 (d), 0.80 (e), 0.90 (f), 0.95 (g), 0.98 (h) from the ULY dataset.



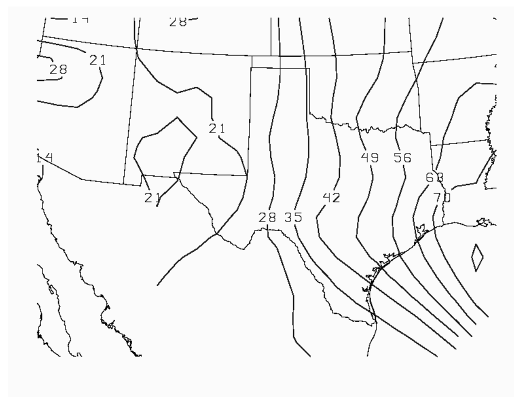
(c) 0.10



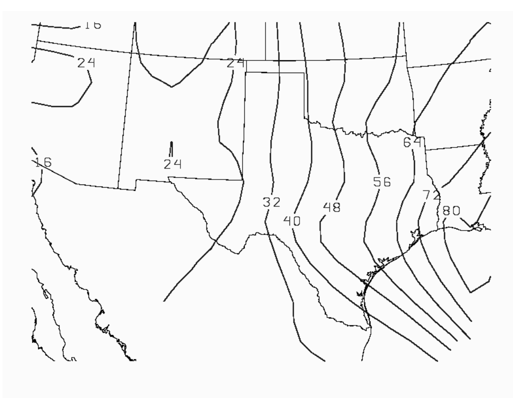
(d) 0.20



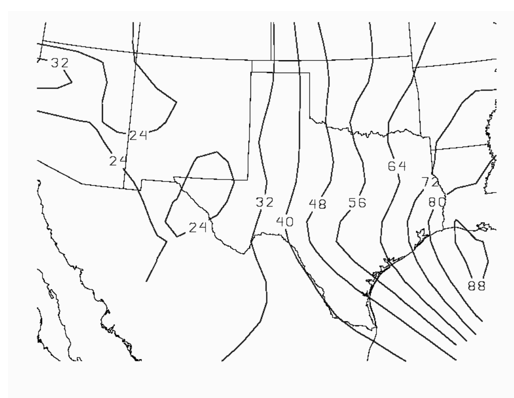
(e) 0.80



(f) 0.90



(g) 0.95



(h) 0.98

FIG 10.1. Continued.

The spatial distribution of the contours for all eight maps in Figure 57. As one would expect, the magnitudes and spatial variances of the precipitation values for all the maps included on Figure 57 are larger in the Eastern half of our domain. An interesting finding with regards to drought is that the areas receiving the least amount of precipitation in a severe drought are outside of New Mexico (Figs 10.1a and Fig. 10.1b). This includes much of the Trans Pecos region (CD 5) in Texas and much of Northern Arizona and Southern Utah. These data would suggest that precipitation is more variable over West Texas than in the state of New Mexico with more frequent droughts. In these regards, Figure 10.1 echoes the claim done on the normalized variance analysis shown in Figure 8.5.

From the COOP climate division distributions, we can create climate division cumulative distribution functions and determine the monthly precipitation values for the aforementioned extreme CDF values (Table 10.1).

TABLE 10.1. Extreme CDF values for New Mexico and Texas 12-month precipitation totals using the CqY dataset.

Climate Division	0.02	0.05	0.10	0.20	0.80	0.90	0.95	0.98
New Mexico 1	6.10	6.95	7.78	8.86	13.97	15.57	16.97	18.64
New Mexico 2	10.40	11.47	12.48	13.79	19.63	21.39	22.92	24.72
New Mexico 3	9.06	10.27	11.44	12.97	20.13	22.35	24.30	26.62
New Mexico 4	8.46	9.47	10.44	11.70	17.45	19.22	20.76	22.58
New Mexico 5	3.30	4.26	5.26	6.67	14.52	17.26	19.76	22.82
New Mexico 6	11.87	12.93	13.93	15.21	20.82	22.48	23.92	25.61
New Mexico 7	7.19	8.30	9.39	10.83	17.72	19.90	21.83	24.13
New Mexico 8	6.12	7.02	7.89	9.05	14.51	16.23	17.74	19.55
Texas 1	11.43	12.80	14.11	15.81	23.59	25.97	28.05	30.51
Texas 2	13.85	15.57	17.21	19.36	29.23	32.26	34.91	38.05
Texas 3	20.51	22.81	25.00	27.83	40.67	44.56	47.95	51.94
Texas 4	30.34	33.31	36.11	39.71	55.71	60.50	64.64	69.48
Texas 5	6.41	7.47	8.50	9.87	16.52	18.64	20.52	22.77
Texas 6	13.12	15.05	16.92	19.39	31.10	34.78	38.02	41.87
Texas 7	17.99	20.50	22.94	26.15	41.21	45.91	50.03	54.92
Texas 8	27.76	30.97	34.04	38.01	56.10	61.61	66.40	72.04
Texas 9	11.44	13.31	15.15	17.61	29.46	33.24	36.58	40.58
Texas 10	13.42	15.30	17.12	19.51	30.75	34.26	37.34	41.00

c. Mann-Whitney Z-Values

To compare trends between different USHCN stations and COOP climate divisions with different precipitation means and variances, it is important to create a dimensionless variable that can indicate extremes in precipitaiton trends. The Z-statistic normalizes the data within a station's time series based on its mean and variance. Higher magnitude Z-statistics that are negative are indicative of drought and higher magnitudes of positive Z-statistics indicate a positive anomaly of precipitation for a given time period.

For precipitation, Z values were calculated using the Mann-Whitney method developed by Mann and Whitney (1947). Mauget (2003) looked at peak periods of the Z-statistic value for both high precipitation ($Z > +1.645$) and drought ($Z < -1.645$) using the Mann-Whitney Z (MWZ) statistic. The MWZ for precipitation at a given time is calculated according to the Mann-Whitney U statistic. This U statistic introduces an element of objectivity and identifies extreme rankings in a sample (Mendenhall et al., 1990).

Depending on the accumulation period of interest, the monthly precipitation totals for a given dataset are divided into two classes of data (Class I and Class II). If one is interested in trends on annual time scales, Class II contains monthly values for one specific year-long period, and the other class (Class I) contains the rest of the monthly values for the entire time series.

For example, in a comparison of the 12-month period in the year 1999 to every other available 12-month period at a given station, Class II contains the twelve monthly values from the year 1999 and Class I contains the rest of the monthly values available in that station's time series. The Mann-Whitney U statistic for each period in Eq. (14) equals the number of Class I members that precede each member of Class II when all the data values are ranked from smallest to largest.. Rank I_i is the rank of the *i*th member of Class I and $\varphi(\text{Rank } I_i, \text{Rank } II_j) = 1$ if $\text{Rank}I_i < \text{Rank } II_j$ and $\varphi(\text{Rank } I_i, \text{Rank } II_j) = 0$ otherwise.

$$U_{II} = \sum_{i=1}^{n_I} \sum_{j=1}^{n_{II}} \varphi\left(Rank\, I_{i,}\ Rank\, II_j\right) \tag{14}$$

For each Mann-Whitney U statistic, one can calculate the MWZ based on a mean (μ_u) and standard deviation (σ_u) of all the U statistics calculated in the time series. In the mean, Eq. (15), and standard deviation, Eq. (16), for the time series of Mann-Whitney , n_I

and n_2 refer to the number of Class I members and Class II respectively. In a study of 12-month accumulation periods, n_2 is equal to twelve. Table 10.2 relates several important CDF probabilities used in this study to MWZ values.

$$\mu_U = \frac{n_1 + n_2}{2} \tag{15}$$

$$\sigma_U = \left(\frac{n_1 n_2 (n_1 + n_2 + 1)}{12} \right)^{1/2} \tag{16}$$

TABLE 10.2. Mann-Whitney Z values matched to several important CDF values.

CDF Value	**0.02**	**0.05**	**0.10**	**0.20**	**0.80**	**0.90**	**0.95**	**0.98**
MWZ Value	-2.325	-1.96	-1.645	-1.28	1.28	1.645	1.96	2.325

d. USHCN Mann-Whitney Z-Value Analysis

Figure 10.2 has running statewide MWZ values for New Mexico and Texas using 12-month period lengths. This analysis has statewide averaged 12-month precipitation totals from each of the available long-term USHCN actual values for each 12-month period. The time stamp for any accumulation period value indicates the last month of in that particular period. Included are values from this dataset in addition to the smoothed, 10-year center-averaged lines for both the Texas and New Mexico time series. The values for the Pacific Decadal Osciallation (PDO) and the El-Niño/Southern Oscillation (ENSO) indices are also includeed, two cycles for which relationships to precipitation are examined.

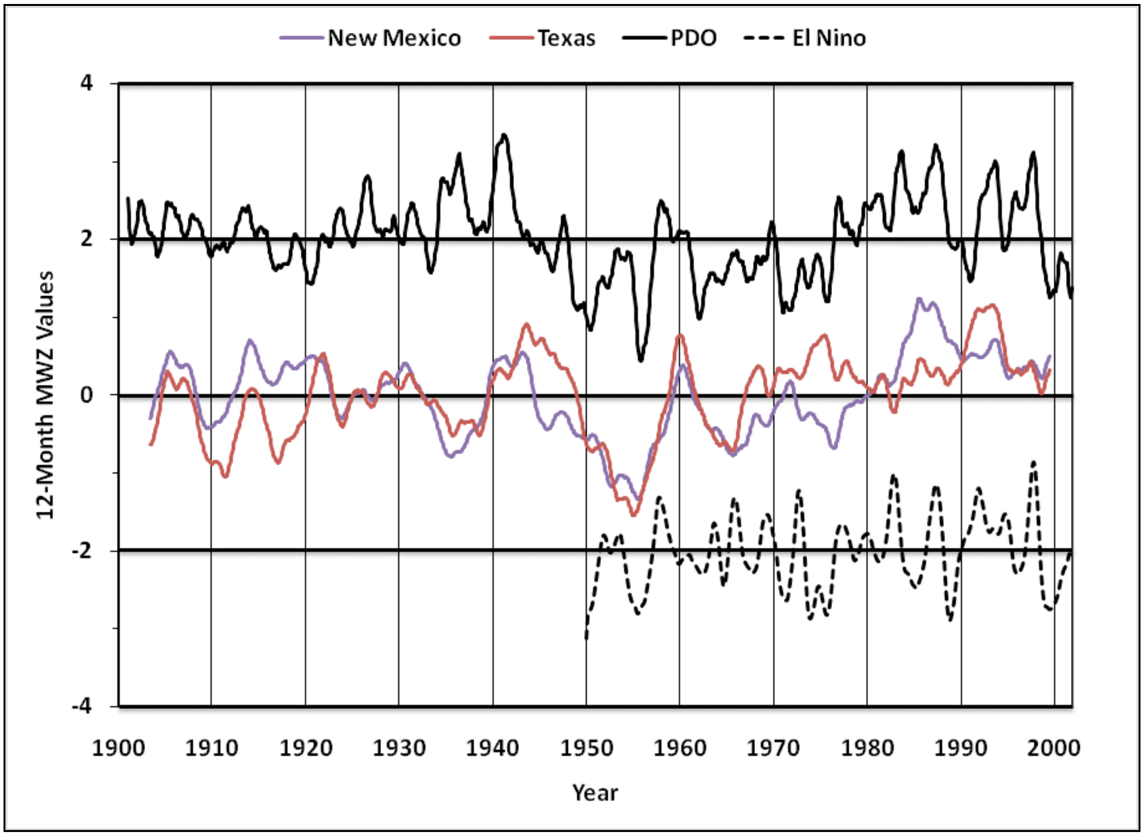


FIG 10.2. Time series of USHCN 12-month precipitation MWZ values for New Mexico (blue) and Texas (red) statewide averages using the ULY dataset. Also included are the values for the ENSO (solid black) and PDO (dotted black) indices, each with its own axis, separate from the central axis for the MWZ values.

The MWZ values for precipitation can show both trends over time and the spatial distribution of precipitation at specific points in time. Of most interest are months with high magnitudes of MWZ values, both positive and negative. Figure 10.2 shows that the long-term precipitation trend in the 20th century can be broken down into three parts. The most notable features are the minima for both New Mexico and Texas in the 1950s. This period of drought in both states is preceded by a noisy and overall indiscernible pattern of precipitation MWZ values in New Mexico and Texas through the first half of the 20th century. The drought of the 1950s is followed by increasing precipitation through the end of the 20th century in both New Mexico and Texas.

The long-term trend in precipitation appears to be leveling off at the end of the 20th century but that brings into question whether this is the beginning of a different long-term trend or a slight fluctuation in the increasing trend. The 12-month running averages of precipitation totals in New Mexico and Texas would suggest a downward trend, but a lack of long-term data beyond this period brings that conclusion into question.

The 12-month accumulation time series (Fig. 10.2) highlights separate peaks in precipitation toward the end of the 20th century for Texas and for New Mexico. The donward trend in statewide averages for New Mexico precipitation begins after a particularly wet period from 1986-1987, in which the El Niño-Southern Oscillation (ENSO) was in a positive phase, better known as an El Niño period. The peak in Texas occurred in the middle 1990s and also coincided with El Niño conditions.

Another cycle that has influence on global weather is the Pacific Decadal Oscillation (PDO), which deals with Pacific sea surface temperatures (SST) north of 20°N, whereas ENSO is an index dealing with SSTs near the Equator. Positive phases of both ENSO and PDO, typically though not always coinciding, generally spell wetter than normal conditions in New Mexico and Texas, whereas the negative phases are correlated with drier than normal conditions. Table 10.3 (Liles 2003) describes the differences in annual precipitation between positive and negative cycles of these two indices for New Mexico Climate divisions.

TABLE 10.3. Relationships between in phase ENSO/PDO episodes and annual precipitation for New Mexico climate divisions (Liles 2003).

	Div 1	**Div 2**	**Div 3**	**Div 4**	**Div 5**	**Div 6**	**Div 7**	**Div 8**
El Niño and PDO+	14.68	18.95	20.82	16.87	12.73	21.06	19.51	13.69
La Niña and PDO-	7.50	11.95	12.78	9.57	7.12	13.26	11.40	8.00

e. COOP Mann-Whitney Z-Value Analysis

The time-series of MWZ shown in Figure 10.2 can be replicated for the individual climate divisions to show the spatial differences in precipitation trends over the past century. All 33 climate divisions in this study are included in the analysis and are grouped into three charts based on the regions denoted in Figure 7.3. The first graph (Fig. 10.3) includes the show the time series for the six East Texas region climate divisions in Texas, the second graph (Fig. 10.4) contains the remaining four Texas CDs in West Texas, and all eight climate divisions in New Mexico are displayed (Fig. 10.5) on the third time series plot.

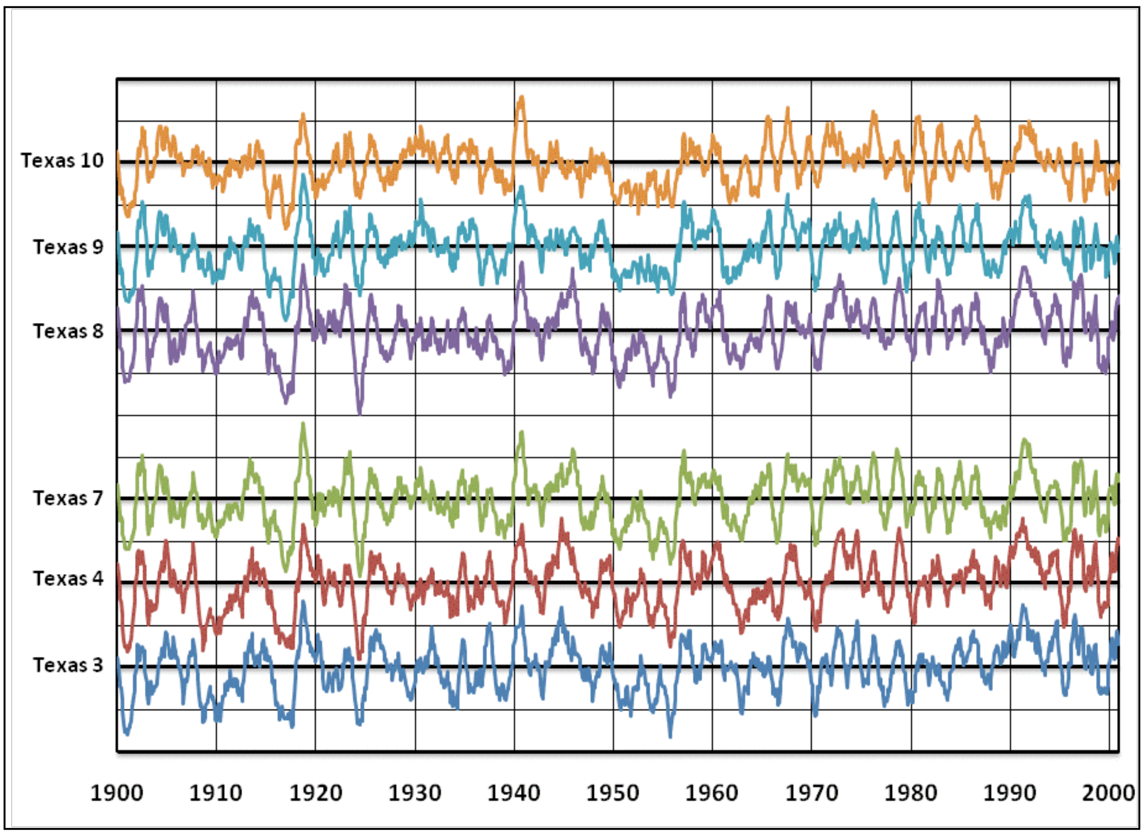


FIG 10.3. Time series of 12-month running precipitation MWZ values for the Texas climate divisions in the East Texas region using the CqY dataset. Each bold line represents a climate division averaged MWZ value of zero and each horizontal line an increment of two.

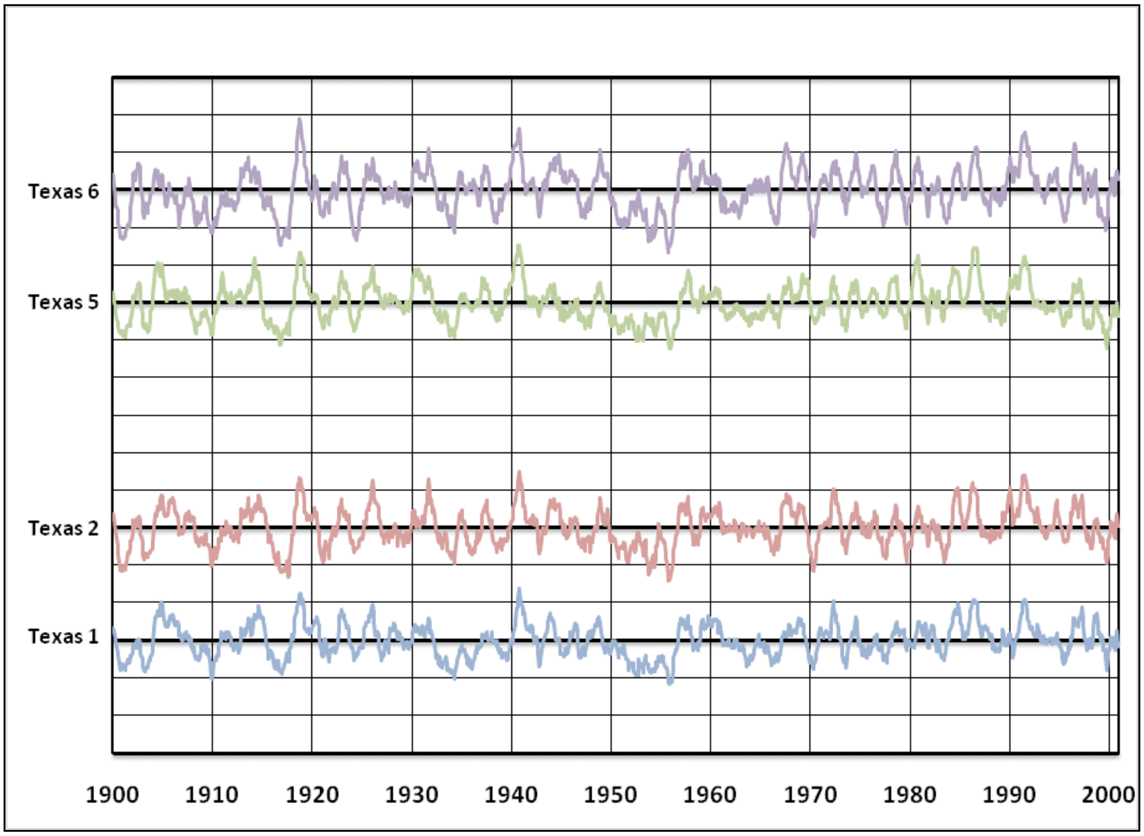


FIG 10.4. Time series of 12-month running precipitation MWZ values for the Texas climate divisions in the West Texas region using the CqY dataset. Each bold line represents a climate division averaged MWZ value of zero and each horizontal line an increment of two.

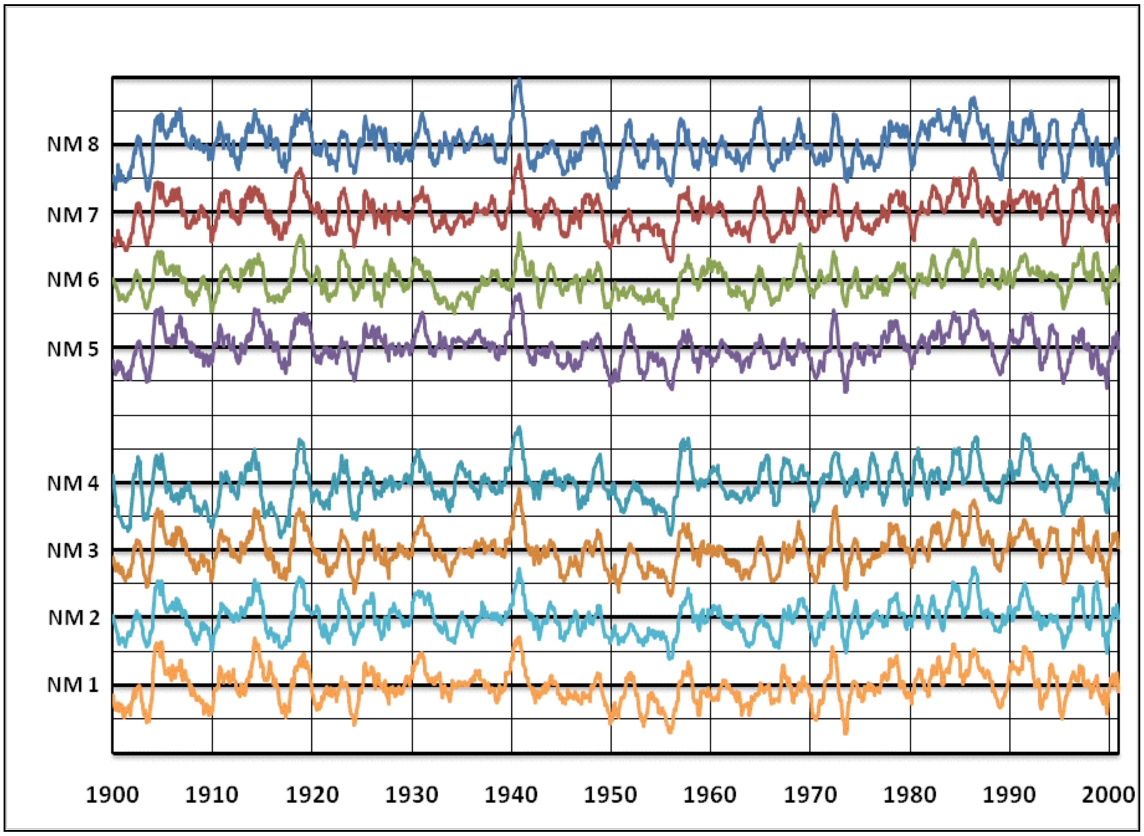


FIG 10.5. Time series of 12-month running precipitation MWZ values for the New Mexico climate divisions in the New Mexico region using the CqY dataset. Each bold line represents a climate division averaged MWZ value of zero and each horizontal line an increment of two.

The figures in this section (Figs. 10.3-10.5) all show roughly the same pattern when the short-term noise is eliminated. The basic pattern is a steady to slightly increasing trend in precipitation over the first half of the 20th century followed by a drastic decrease in precipitation in the 1950s. Following the long-term drought lasting through most of the 1950s, there was an increasing trend in precipitation through most of the latter half of the 20th century.

However, there is some evidence of a downturn in MWZ precipitation values for the last few years in most of the individual climate division time series datasets. This signal is most prominent in West Texas and New Mexico (Figs. 10.4 and 10.45) and not in East Texas (Fig. 10.3). This result would be indicative of a shift towards drier conditions in areas already relatively dry and wetter conditions in areas already with relatively wet overall conditions. However, any conclusions implicating a long-term shift in precipitation trend need several more years of data to conclude that is not in fact just a short-term fluctuation in the overall precipitation pattern.

f. COOP Climate Division Decadal Precipitation Averages

There are several ways to measure precipitation trends on decadal scales, with perhaps the simplest measure being the annual average precipitation by decade. Table 10.4 contains the decadal annual precipitation averages for each of the 18 New Mexico and Texas COOP climate divisions using the CqY dataset.

TABLE 10.4. Annual average precipitation by decade for division averaged precipitation using the CqY dataset.

	1900s	1910s	1920s	1930s	1940s	1950s	1960s	1970s	1980s	1990s
New Mexico 1	12.73	12.69	11.59	11.66	10.83	11.20	11.48	11.78	14.60	12.94
New Mexico 2	16.54	18.40	16.85	15.70	16.85	15.00	15.54	15.38	18.45	18.22
New Mexico 3	15.54	17.60	16.88	13.52	17.39	14.80	15.21	15.58	18.66	18.50
New Mexico 4	14.76	16.60	14.45	14.62	12.83	12.78	13.10	13.46	15.94	14.48
New Mexico 5	11.50	12.72	10.27	12.75	11.60	9.44	10.16	14.40	12.20	12.03
New Mexico 6	16.22	19.33	17.30	17.74	17.85	15.69	16.61	17.57	20.74	18.77
New Mexico 7	13.55	14.90	13.87	13.11	14.35	11.65	12.24	14.35	15.81	15.18
New Mexico 8	12.41	13.79	11.78	12.19	10.91	10.37	11.47	11.93	14.50	13.34
Texas 1	18.64	20.32	19.92	16.23	21.03	16.84	17.69	18.65	20.47	20.14
Texas 2	22.92	24.14	23.50	22.72	24.67	21.07	23.25	24.24	25.42	25.22
Texas 3	31.44	33.30	32.49	33.31	35.74	31.04	33.51	33.59	36.15	37.82
Texas 4	44.63	44.82	44.80	45.27	48.94	43.67	42.90	47.78	48.23	51.54
Texas 5	14.99	13.19	13.34	13.03	13.54	10.51	11.74	14.44	15.04	12.63
Texas 6	22.46	25.61	24.40	26.37	25.84	22.72	24.29	27.68	26.21	27.41
Texas 7	31.34	32.76	32.37	33.30	33.71	31.56	33.17	37.21	32.84	37.52
Texas 8	45.95	46.10	44.64	45.89	49.82	42.73	44.21	50.77	47.51	53.68
Texas 9	22.16	23.12	22.58	23.29	24.22	21.78	23.17	26.81	23.24	24.27
Texas 10	-1.00	22.84	24.38	24.16	23.74	22.12	26.54	28.46	23.99	25.23

g. COOP Climate Division Decadal Precipitation Extremes

Another measure of precipitation was developed to track changes in extreme precipitation on decadal time scales. This methodology uses the climate division averaged COOP time series data and produces a statistic based on MWZ monthly precipitation values. For each decade, this statistic determines the percentage of MWZ values below the precipitation value representing a CDF of 0.10 in Table 10.5. Similarly, each climate division has a decadal percentage of MWZ values above the precipitation value representing a CDF of 0.90 (Table 10.6). These percentages for the 18 climate divisions in New Mexico and Texas are calculated using the CqY dataset.

TABLE 10.5. Percentage of MWZ precipitation values for the 18 New Mexico and Texas climate divisions below the CDF precipitation value of 0.10. Data are organized by decade and used the CqY dataset.

	1900s	1910s	1920s	1930s	1940s	1950s	1960s	1970s	1980s	1990s
New Mexico 1	14.17	0.83	0.83	0.00	7.50	15.00	0.00	4.17	4.17	6.67
New Mexico 2	10.00	0.00	1.67	0.00	5.83	11.67	2.50	0.00	0.00	3.33
New Mexico 3	1.67	0.83	0.00	3.33	0.00	6.67	0.83	0.00	0.00	0.00
New Mexico 4	8.33	0.00	2.50	0.00	2.50	14.17	0.00	7.50	0.00	5.00
New Mexico 5	24.17	25.00	6.67	0.83	0.00	15.83	0.00	0.83	0.00	2.50
New Mexico 6	6.67	1.67	4.17	0.00	2.50	14.17	6.67	5.83	0.00	2.50
New Mexico 7	2.50	4.17	0.00	0.00	0.00	5.83	5.00	2.50	0.00	1.67
New Mexico 8	10.00	0.83	5.00	0.00	1.67	25.00	1.67	11.67	0.00	1.67
Texas 1	2.50	4.17	0.00	4.17	0.00	14.17	0.00	0.00	0.00	0.00
Texas 2	7.50	12.50	1.67	0.83	0.00	16.67	0.00	4.17	2.50	0.83
Texas 3	18.33	23.33	7.50	4.17	0.00	21.67	4.17	5.00	0.00	0.00
Texas 4	19.17	25.00	8.33	2.50	0.00	16.67	6.67	5.00	1.67	2.50
Texas 5	5.00	7.50	0.00	3.33	0.00	16.67	0.00	0.00	0.00	3.33
Texas 6	15.00	16.67	6.67	5.00	0.00	24.17	2.50	6.67	0.83	3.33
Texas 7	8.33	15.83	9.17	1.67	0.83	25.83	6.67	4.17	0.83	3.33
Texas 8	10.00	20.83	9.17	5.83	2.50	23.33	3.33	3.33	2.50	5.83
Texas 9	9.17	15.83	5.83	3.33	0.00	12.50	2.50	5.00	0.00	0.00
Texas 10	8.33	15.83	0.83	0.00	0.00	15.00	3.33	0.00	1.67	2.50

TABLE 10.6. Percentage of MWZ precipitation values for the 18 New Mexico and Texas climate divisions above the CDF precipitation value of 0.90. Data are organized by decade and used the CqY dataset.

	1900s	1910s	1920s	1930s	1940s	1950s	1960s	1970s	1980s	1990s
New Mexico 1	7.50	6.67	0.83	2.50	10.83	0.00	3.33	2.50	15.00	4.17
New Mexico 2	5.83	9.17	0.00	0.00	9.17	0.00	0.00	0.83	12.50	2.50
New Mexico 3	1.67	8.33	1.67	0.00	5.00	0.00	2.50	0.83	11.67	0.83
New Mexico 4	9.17	13.33	0.00	3.33	11.67	0.00	0.00	4.17	11.67	6.67
New Mexico 5	0.83	9.17	0.00	2.50	12.50	10.00	0.83	7.50	15.00	11.67
New Mexico 6	7.50	15.00	0.00	1.67	10.83	0.00	0.83	5.00	16.67	5.00
New Mexico 7	7.50	12.50	0.83	0.00	9.17	0.83	0.00	1.67	15.83	8.33
New Mexico 8	9.17	9.17	0.00	5.83	10.83	0.00	0.00	6.67	17.50	6.67
Texas 1	3.33	7.50	3.33	0.00	6.67	0.00	0.00	1.67	10.00	5.00
Texas 2	0.83	7.50	4.17	2.50	5.83	0.00	1.67	3.33	13.33	10.00
Texas 3	1.67	6.67	0.83	5.83	16.67	1.67	5.00	6.67	2.50	20.83
Texas 4	6.67	6.67	0.00	0.00	20.00	3.33	8.33	19.17	0.83	28.33
Texas 5	7.50	10.83	1.67	3.33	9.17	1.67	0.00	0.00	15.83	8.33
Texas 6	0.00	9.17	2.50	1.67	13.33	5.00	9.17	2.50	6.67	15.00
Texas 7	4.17	10.00	5.00	0.00	15.00	2.50	2.50	11.67	0.83	16.67
Texas 8	5.83	8.33	5.83	0.00	16.67	1.67	1.67	15.83	5.00	29.17
Texas 9	5.00	7.50	1.67	1.67	10.83	3.33	7.50	5.83	5.83	8.33
Texas 10	3.33	4.17	0.00	0.83	10.00	0.00	9.17	6.67	11.67	6.67

The analyses done on decadal averages and on the decadal distribution extremes back the other analyses in that the 1950s was the decade with the driest conditions and the 1990s were the driest decade overall. Nine of the ten climate divisions in Texas and six of the eight climate divisions in New Mexico (Table 10.5) had their driest decade from 1951-1960. Drought in this sense is loosely defined as having monthly precipitation totals below the 10th precentile of their climate division's distribution.

Table 10.6 provides overwhelming evidence that the largest number of months with extremely high precipitation occurred in the last two decades of the 20th century. Six of the ten Texas climate divisions count the period from 1991-2000 as its highest average from the ten decades in the 20th century. Seven of the eight climate divisions in New Mexico (Table 10.6) have their wettest decade from 1981-1990 with particularly high 12-month MWZ values (Fig. 10.2) in the period from 1986-1987.

h. Percentage of Moderate and Exceptional Drought Months in COOP Climate Divisions

Further drought analyses were done on the COOP stations in Texas and New Mexico for the entirety of the four 25-year periods and ten 10-year periods in the 20th century. The severity of drought is often thought of as an extension of an area's overall precipitation distribution. The four lower CDF probabilities of interest are shown in Table 10.2, ranging from the 20th percentile to the 2nd. This study characterizes periods with precipitation totals below the 20th percentile of its given distribution as a moderate drought and months below its 2nd percentile as more exceptional droughts, following the guidelines of the United States Drought Monitor (Svoboda et al. 2002).

The following set of analyses focuses on the 20th and 2nd percentile of each COOP station's overall CqY distribution. Each analysis will be a map color-coded according to the percentage of months below each drought threshold. Figures 10.6 and 10.7 each contain four maps of the 25-year periods, the first pertaining to the 20th percentile (Fig. 10.6) and the second for the 2nd percentile (Fig. 10.7). Figures 10.8 (20th percentile) and 10.9 (2nd percentile) each contain maps of the 10-year periods in the 20th century.

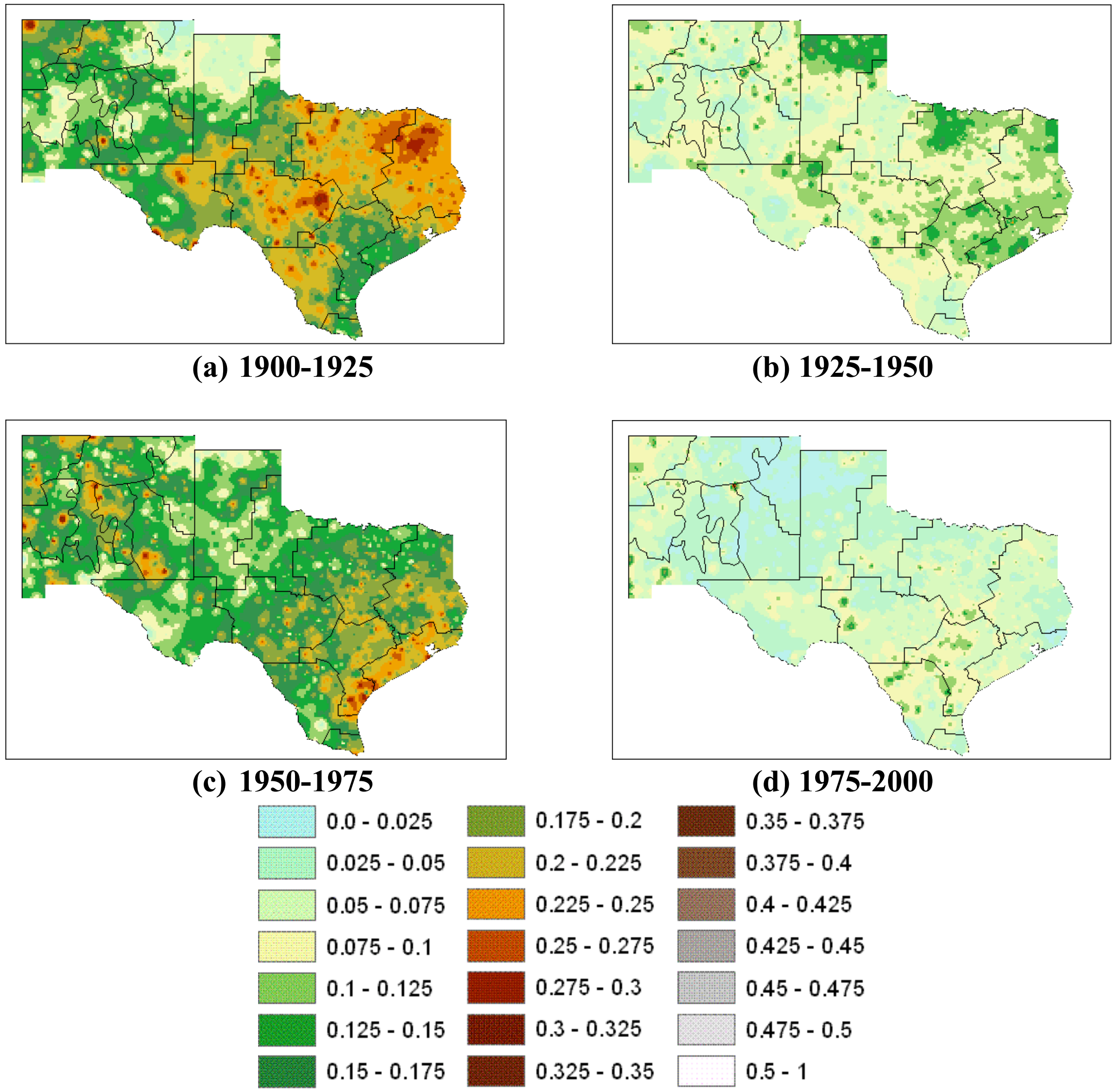


FIG 10.6. COOP climate division maps color-coded according to the percentage of months below the 20th percentile of its given distributions for the periods 1900-1925 (a), 1925-1950 (b), 1950-1975 (c), and 1975-2000 (d) using the CqY dataset. The legend denotes the fractional percentage for the colors on the maps.

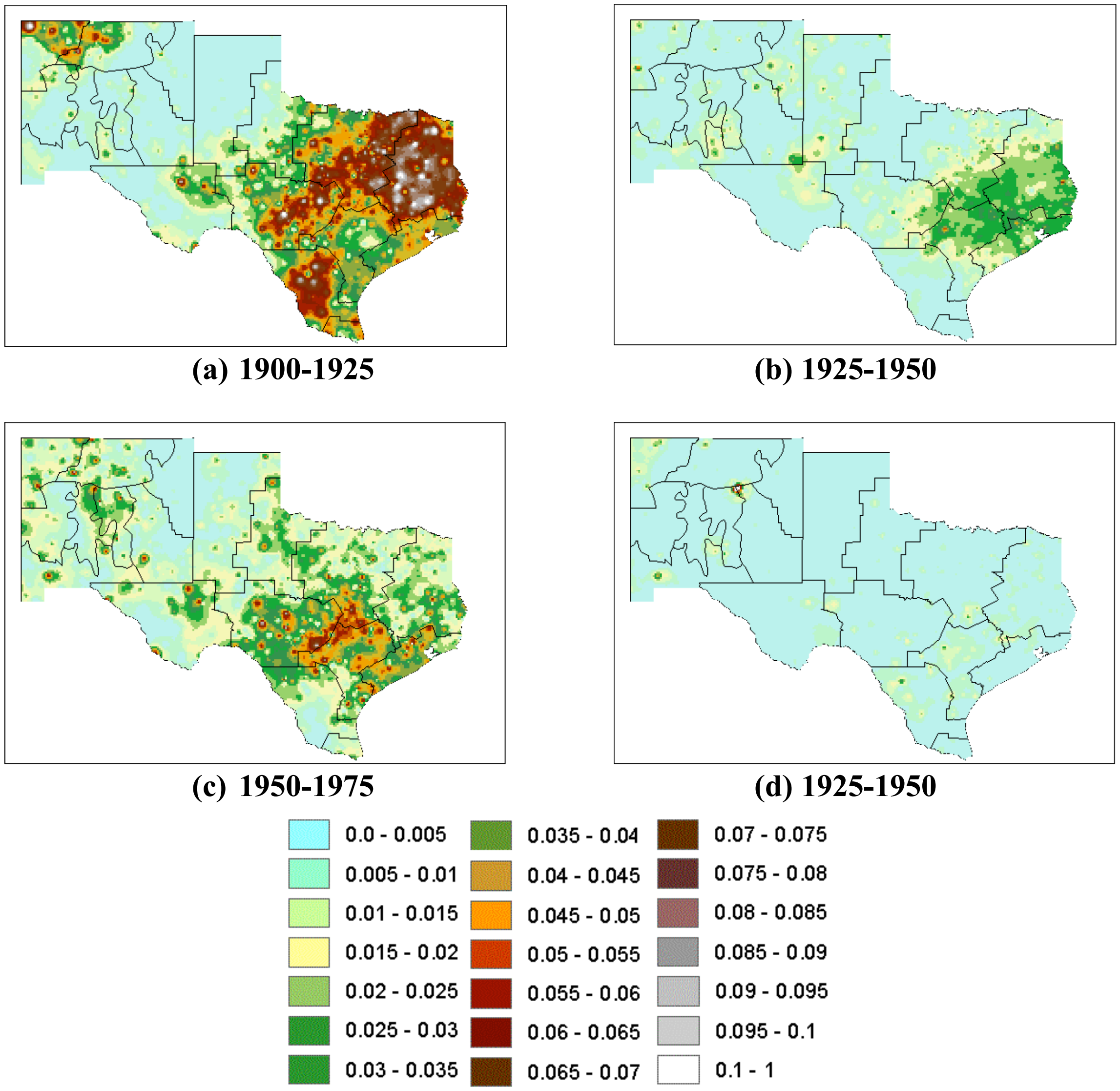
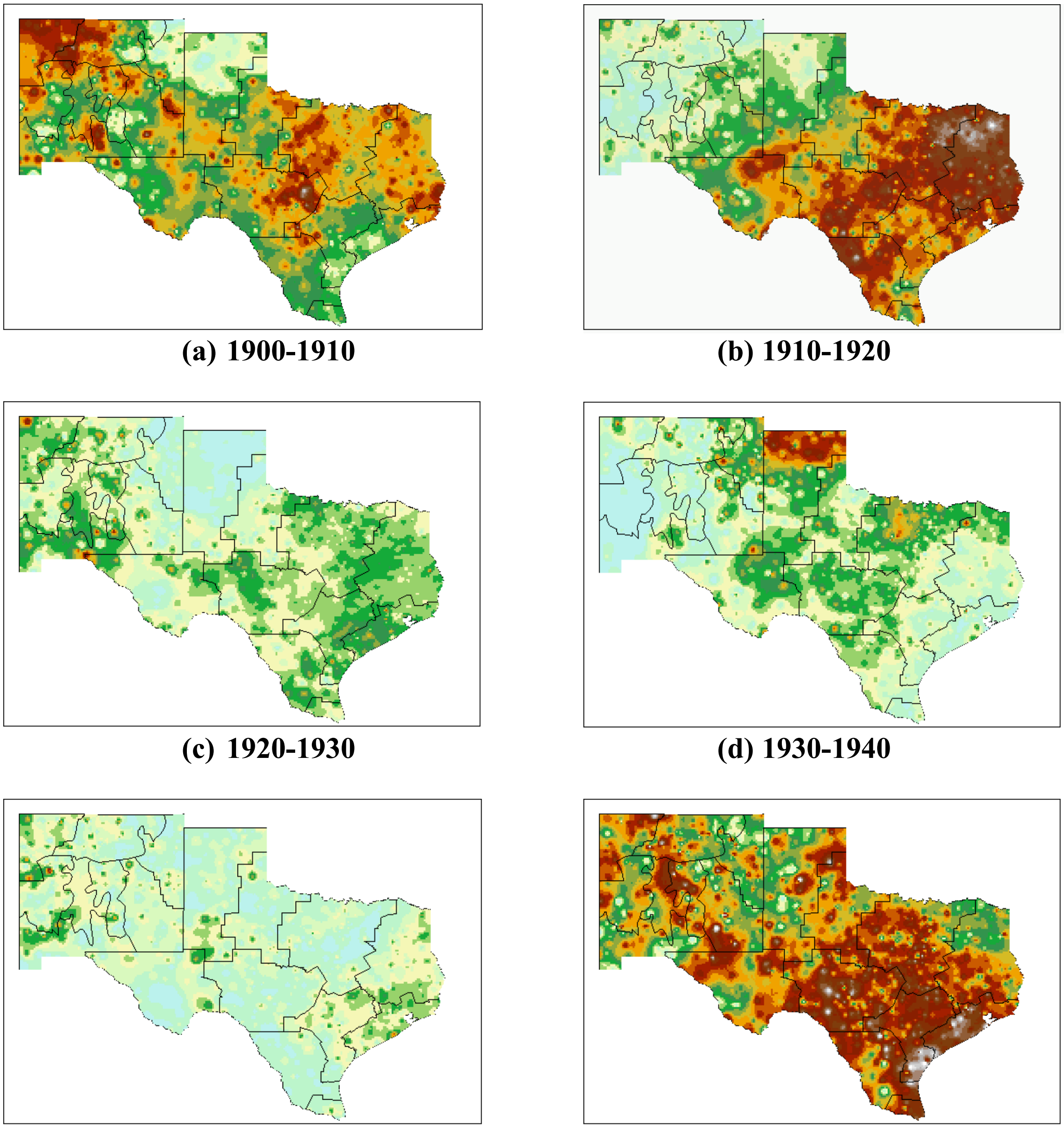


FIG 10.7. COOP climate division maps color-coded according to the percentage of months below the 2nd percentile of its given distributions for the periods 1900-1925 (a), 1925-1950 (b), 1950-1975 (c), and 1975-2000 (d) using the CqY dataset. The legend denotes the fractional percentage for the colors on the maps.



(a) 1900-1910 (b) 1910-1920

(c) 1920-1930 (d) 1930-1940

(e) 1940-1950 (f) 1950-1960

FIG 10.8. COOP climate division maps color-coded according to the percentage of months below the 20th percentile of its given distributions for the periods 1900-1910 (a), 1910-1920 (b), 1920-1930 (c), and 1930-1940 (d), 1940-1950 (e), 1950-1960 (f), 1960-1970 (g), and 1970-1980 (h), 1980-1990 (i), 1990-2000 (j) using the CqY dataset. The legend denotes the fractional percentage for the colors on the maps.

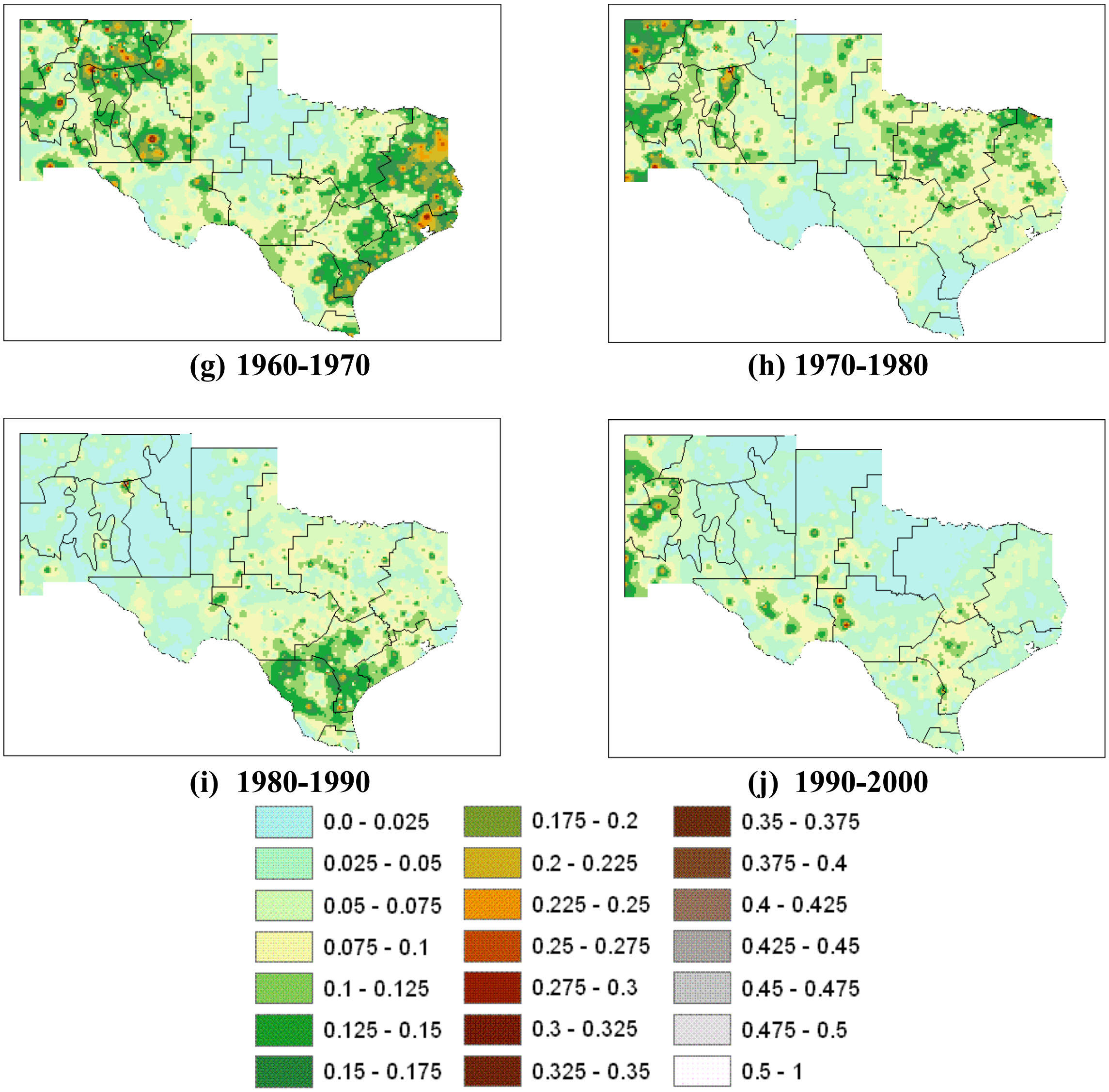
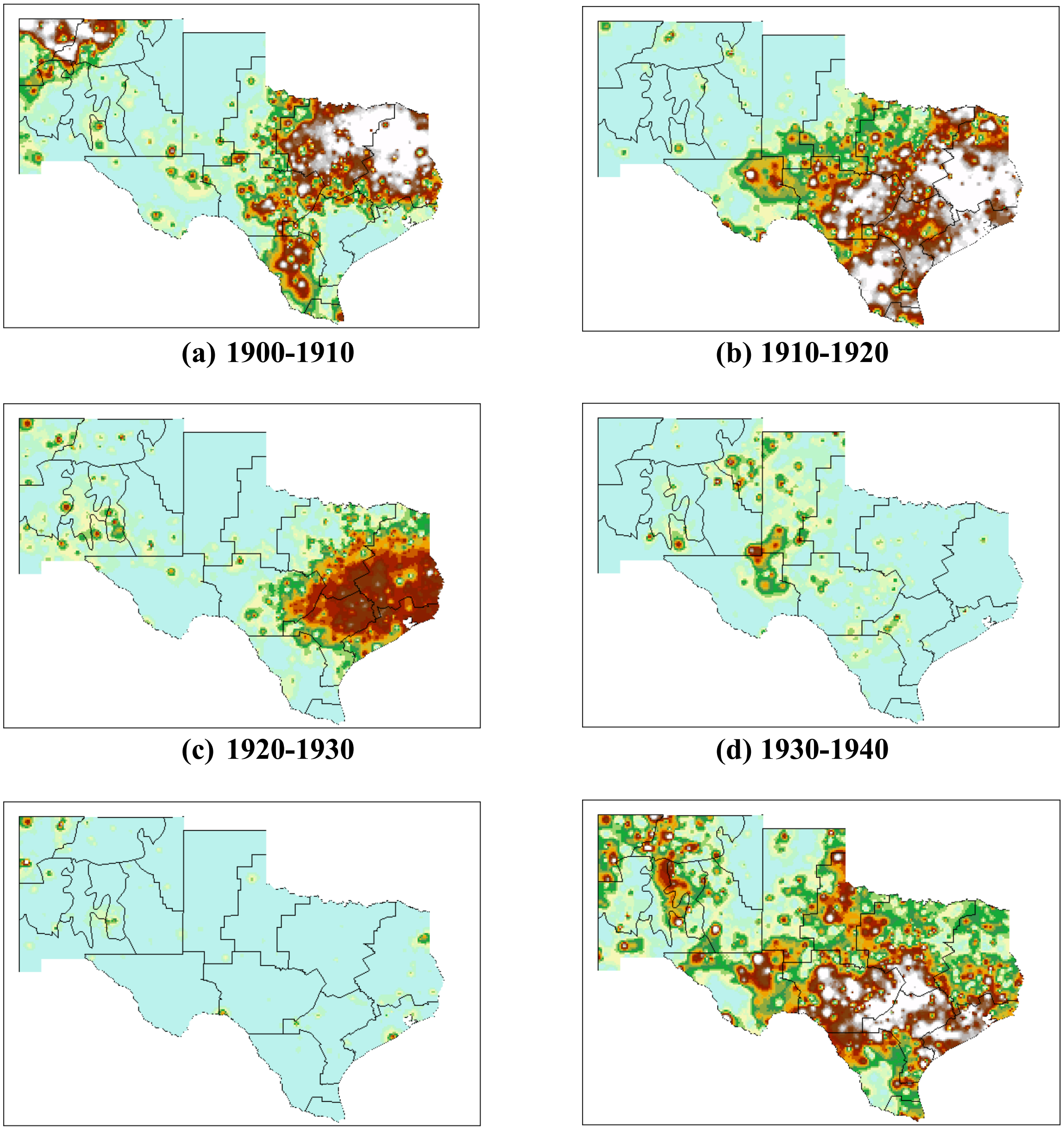


FIG 10.8. Continued.

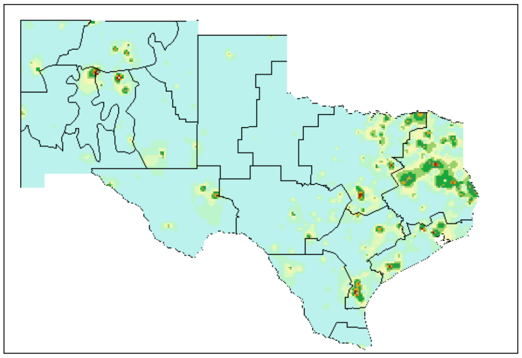


(a) 1900-1910 (b) 1910-1920

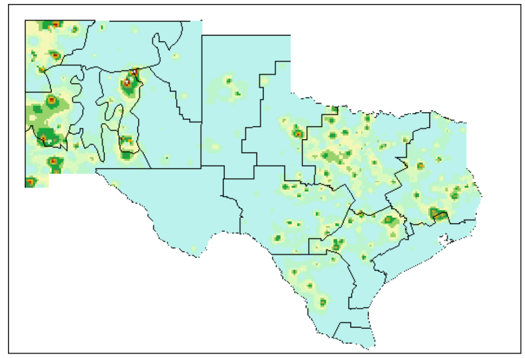
(c) 1920-1930 (d) 1930-1940

(e) 1940-1950 (f) 1950-1960

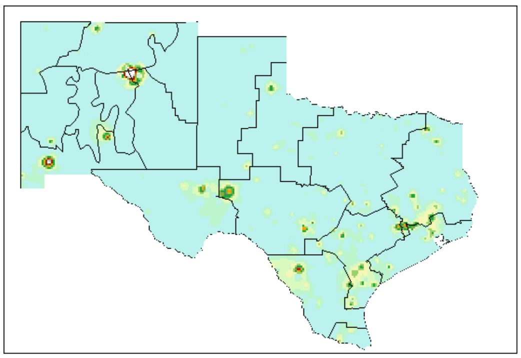
FIG 10.9. COOP climate division maps color-coded according to the percentage of months below the 2nd percentile of its given distributions for the periods 1900-1910 (a), 1910-1920 (b), 1920-1930 (c), and 1930-1940 (d), 1940-1950 (e), 1950-1960 (f), 1960-1970 (g), and 1970-1980 (h), 1980-1990 (i), 1990-2000 (j) using the CqY dataset. The legend denotes the fractional percentage for the colors on the maps.



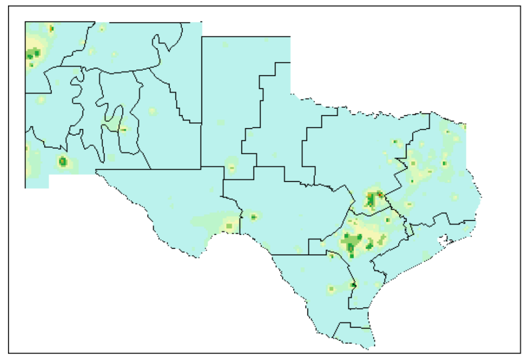
(g) 1960-1970



(h) 1970-1980



(i) 1980-1990



(j) 1990-2000

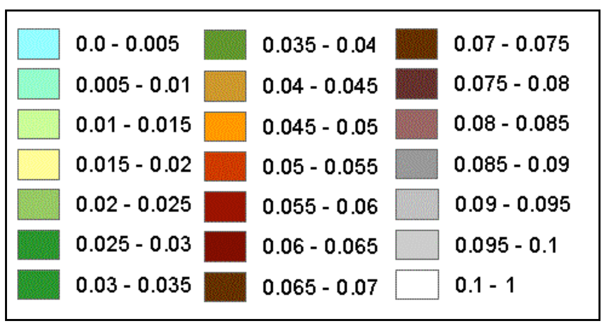


FIG 10.9. Continued.

Figures 10.6 through 10.9 provide an excellent visualization of the spatial differences in moderate (below 20^{th} percentile) and exceptional (below 2^{nd} percentile) droughts. Beginning with Figure 10.6, it is clear to see that the first and third 25-year periods of the 20^{th} century contained the most months with conditions below the 20^{th} percentile of each station's 12-month running time series.

The years 1900-1925 (Fig 10.6a) were most affected by drought in Northeast and Central Texas, with the rest of Texas and New Mexico experiencing more months with moderate drought in the 1950-1975 period. Judging by Figure 10.7, the most frequent exceptional droughts in New Mexico and Texas were in Northeast Texas toward the beginning of the 20^{th} century. Figure 10.3 shows the largest magnitude negative values for Texas CDs 3 and 4 occurred in 1917.

The maps for percentage of moderate drought months by decade (Fig. 10.8) show the 10-year period of 1950-1960 to have the most drought months for each region in New Mexico and Texas with the exception of the panhandle of Texas and far East Texas (Fig. 10.8f). The 1930-1940 time period was the driest in the panhandle region in association with the Dust Bowl that was prevalent across the Great Plains. The 1910-1920 time period was most responsible for the high percentage of droughts in the first quarter of the 20^{th} century in East Texas.

According to Figs. 10.9a and 10.9b, the greatest frequency of exceptional droughts occurred in the first two decades of the 20^{th} century and was prevalent over East Texas. The drought of the 1950s was most severe in the present-day 1-35 corridor from Austin to San Antonio. With the exception of the period 1950-1960, the number of severe droughts from 1930 through 2000 was extremely low when compared to the numbers from 1900-1930.

The next section deals with the more specfic durations of drought, breaking down the exact time periods of climate-division drought for the four different drought thresholds of interest. The section following this next penultimate section adjust the USHCN state averaged and COOP climate division averaged variance-adjusted time series for the century-long trend. Would a changing mean of precipitation cause the inordinate amounts of moderate to severe droughts in the first part of the 20th century and a relatively low number toward the end of the 20th century? The maps shown for the decadal percentages of moderate and severe droughts will be recreated after adjusting the time series of variance-adjusted precipitation.

i. Durations below the Four Drought Thresholds for COOP Climate Division Data

The figures associated with the previous section (Figs. 10.6-10.9) depicting the percentage of months below the two drought threshold cannot tell the entire story of drought throughout the 20th century, nor can they tell the entire story for the shorter periods of time. The goal of this section is to provide a basic visualization of the specific period of droughts below each of our four drought thresholds.

For each of the four thresholds, there is a straight line that will have patches of color denoting the intensity of drought for certain periods These patches will show up if a climate division averaged 12-month precipitation value from the CqY dataset was below the 20th percentile. The four 25-year quarters of the 20th century will be investigated (Figs. 10.10-10.13) rather than the 20th century as a whole in order to get a better resolution of the different droughts and intensities of these droughts.

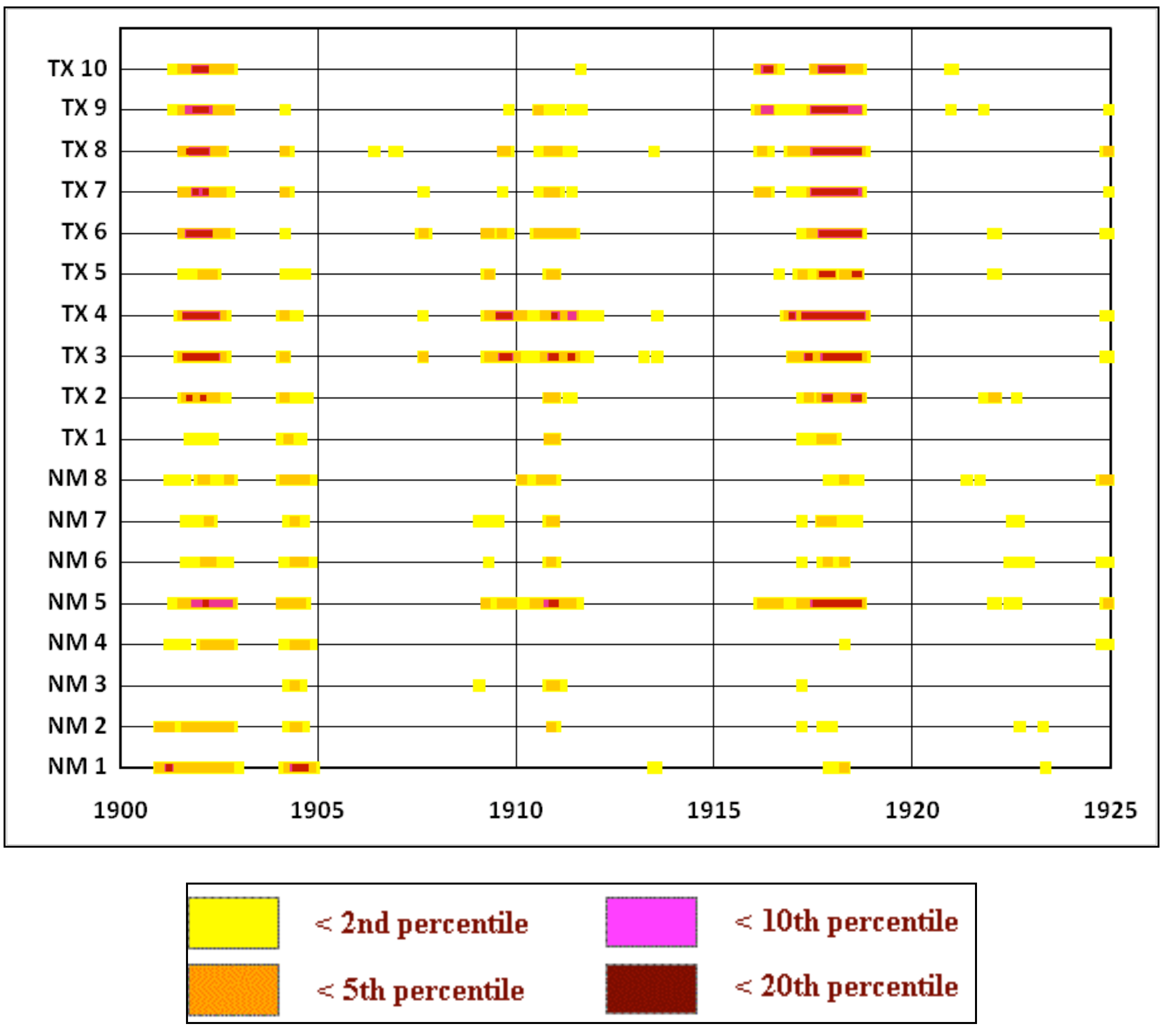


FIG 10.10. Periods in the time frame 1900-1925 when 12-month running climate division averaged values, from the CqY dataset were below the thresholds specified by the colors.

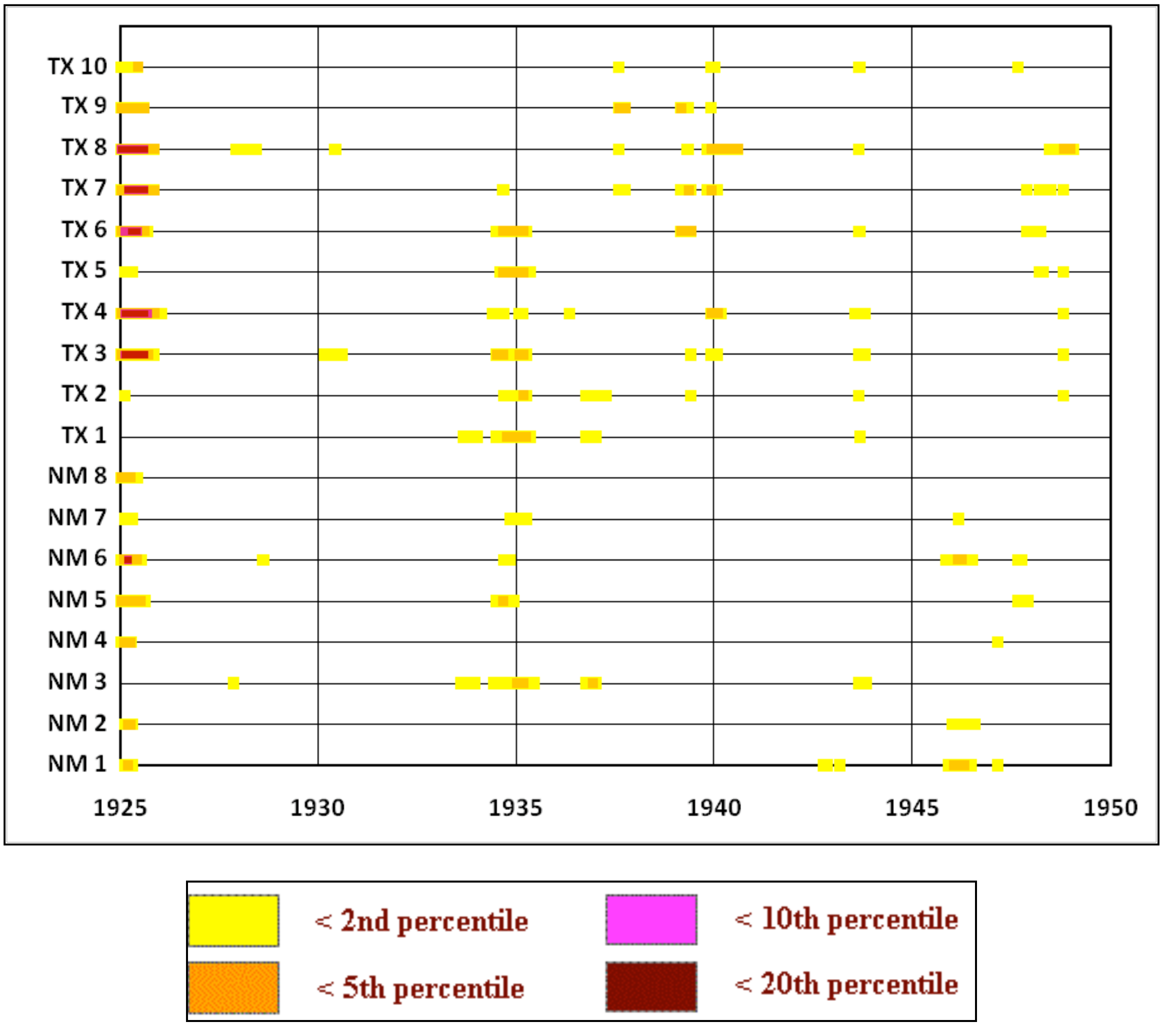


FIG 10.11. Periods in the time frame 1925-1950 when 12-month running climate division averaged values, from the CqY dataset were below the thresholds specified by the colors.

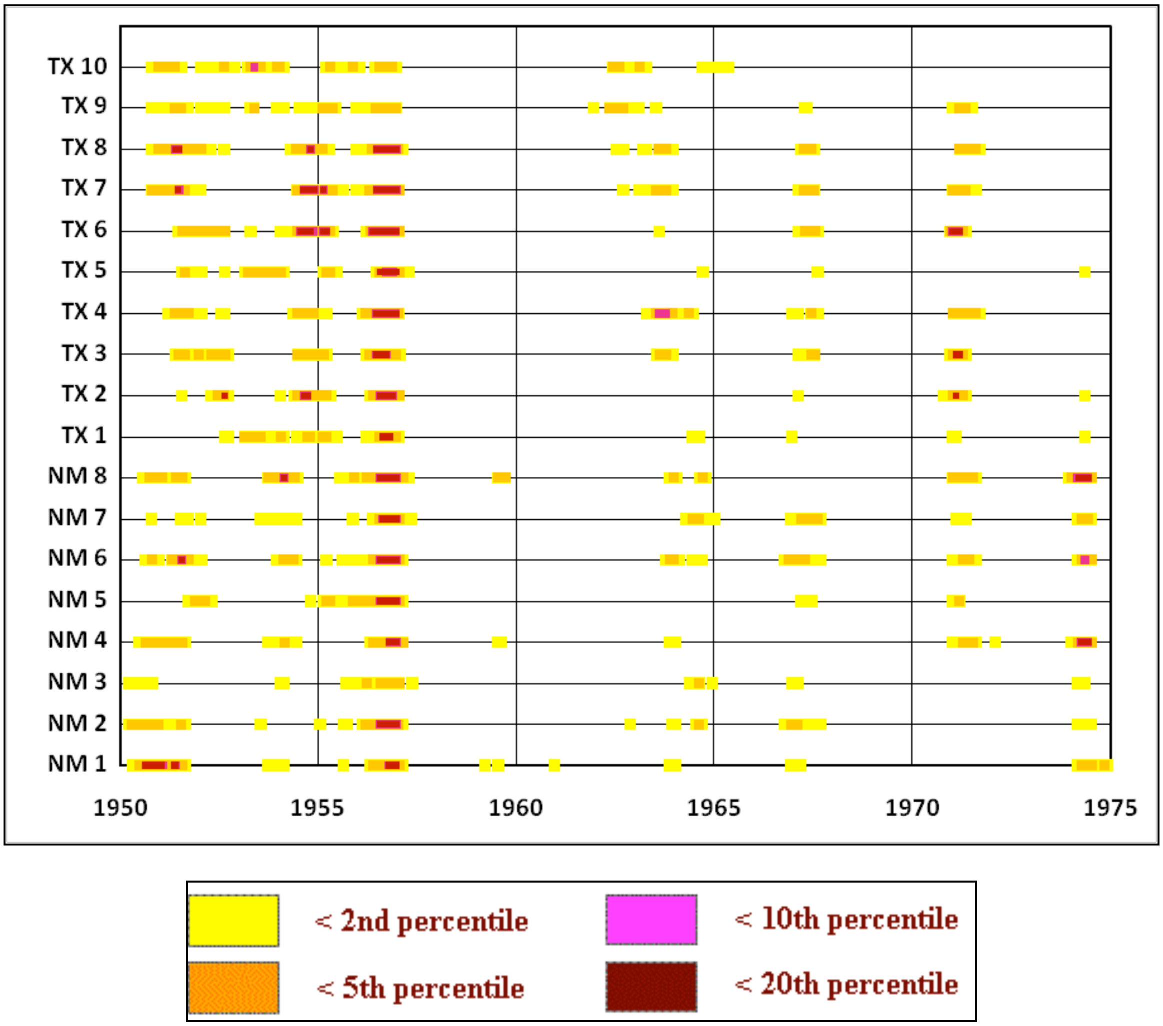


FIG 10.12. Periods in the time frame 1950-1975 when 12-month running climate division averaged values, from the CqY dataset were below the thresholds specified by the colors.

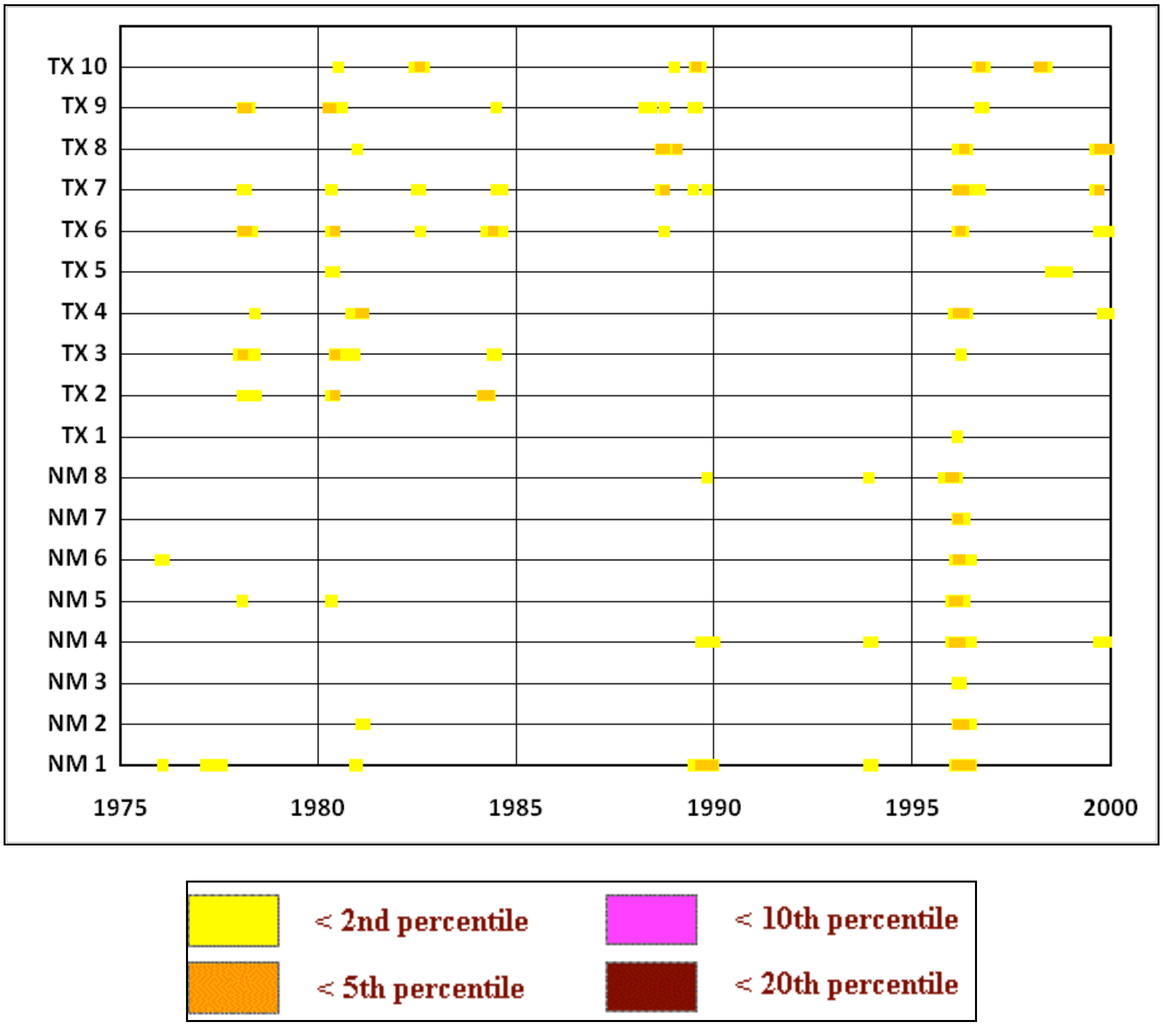


FIG 10.13. Periods in the time frame 1975-2000 when 12-month running climate division averaged values, from the CqY dataset were below the thresholds specified by the colors.

The analyses of drought duration for different thresholds based on 12-month running precipitation totals make it very clear when drought affects the different climate divisions. Every climate division in New Mexico and Texas experiences at least a moderate drought through most of the 1950s. These analyses also back up the previous conclusion that the 1980s were a wet decade throughout the entire state of New Mexico. The state of Texas, based on the drought definitions based on percentiles of distributions, had very few major drought episodes from the drought period of the 1950s through the end of the 20th century. Exceptional droughts were frequent from 1900-1930 across Texas but none of these periods were sustained as long as the drought in the 1950s.

The lengths of droughts appeared to be longest in the first 25-year period and during the long-term drought of the 1950s. Generally speaking, the periods of drought toward the end of the 20th century were much shorter in duration than those during the first half of the century. Also, the severity of droughts has lessened toward the end of the 20th century when compared to the earlier drought periods. There were only four short during the last 25-year period in which 12-month precipitation totals fell below the 5th percentile (red on Fig.10.13) and at no time in any climate division was there a 12-month exceptional drought.

j. Adjustment of Precipitation Values for Century-Long Precipitation Trends

The analyses of precipitation trend show that for the most part, precipitation is increasing on all time scales. The literature review on drought and precipitation trends suggests most of this occurs in extreme precipitation events. Whatever the case, mean precipitation is not a static quantity so it is useful to adjust the precipitation distribution to account for these changes over time. This last section will focus on analyses of precipitation time series using the trends in ULY station data (Fig. 9.1) and CqY climate division averaged data (Table 9.1), which formed the UTY dataset for the USHCN long-term stations and the CTY dataset based on modification of the CqY dataset.

The first analysis adjusts the 12-month precipitation totals according to the USHCN trends, creating a new time series plot for the statewide averages from Texas and New Mexico (Fig. 10.14). The second group of analyses (Figs. 10.15-10.17) is similar to the MWZ analyses in Figures 10.2 through 10.5 and again, adjusts the 12-month MWZ values to adjust for the 20th century trends in precipitation for each COOP station.

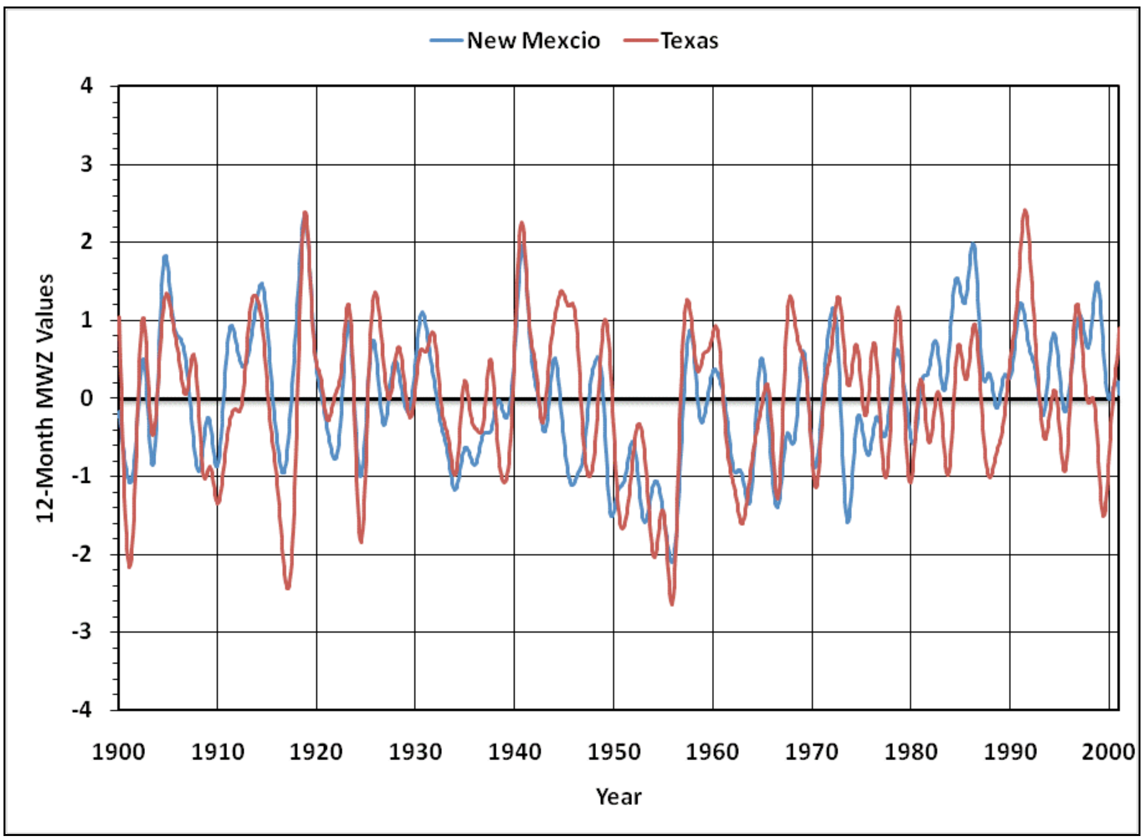


FIG 10.14. Time series of 12-month running precipitation MWZ values. These values are statewide averages for both New Mexico and Texas using actual the UTY dataset.

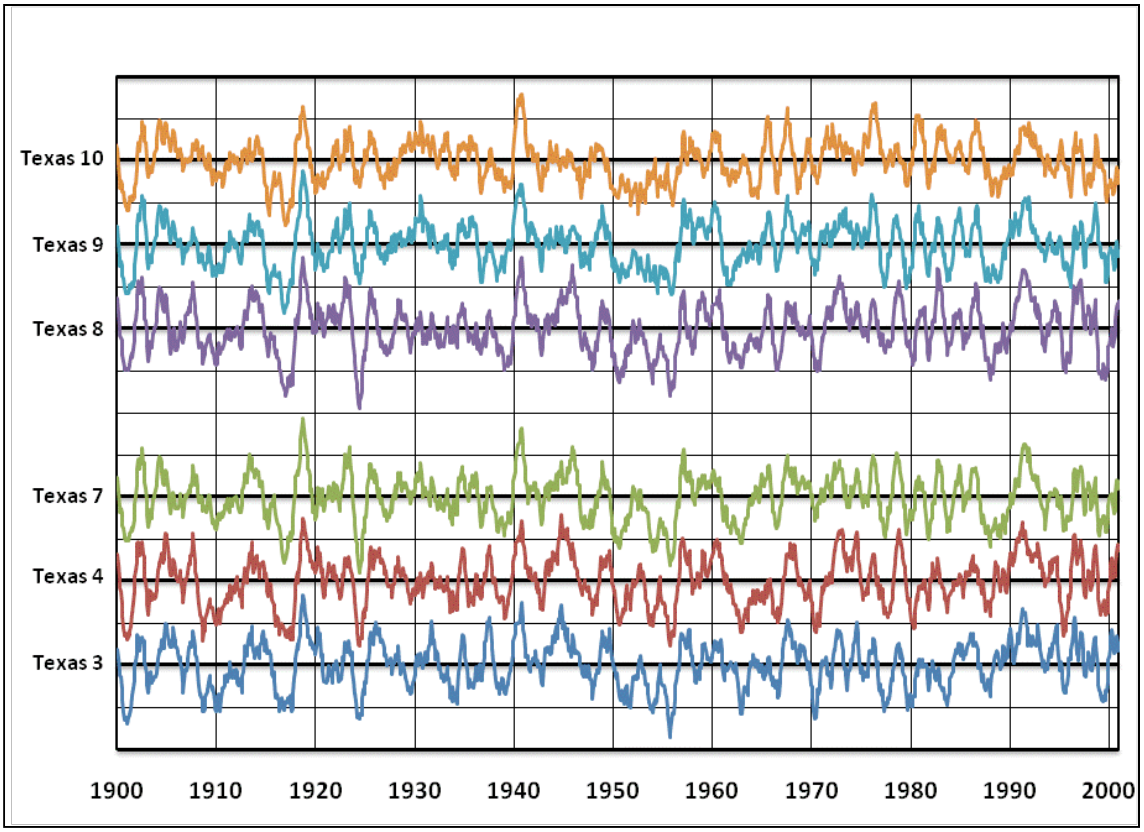


FIG 10.15. Time series of climate division averaged 12-month running precipitation MWZ values for the East Texas region using the CTY dataset. Each bold line represents a climate division averaged MWZ value of zero and each horizontal line an increment of two.

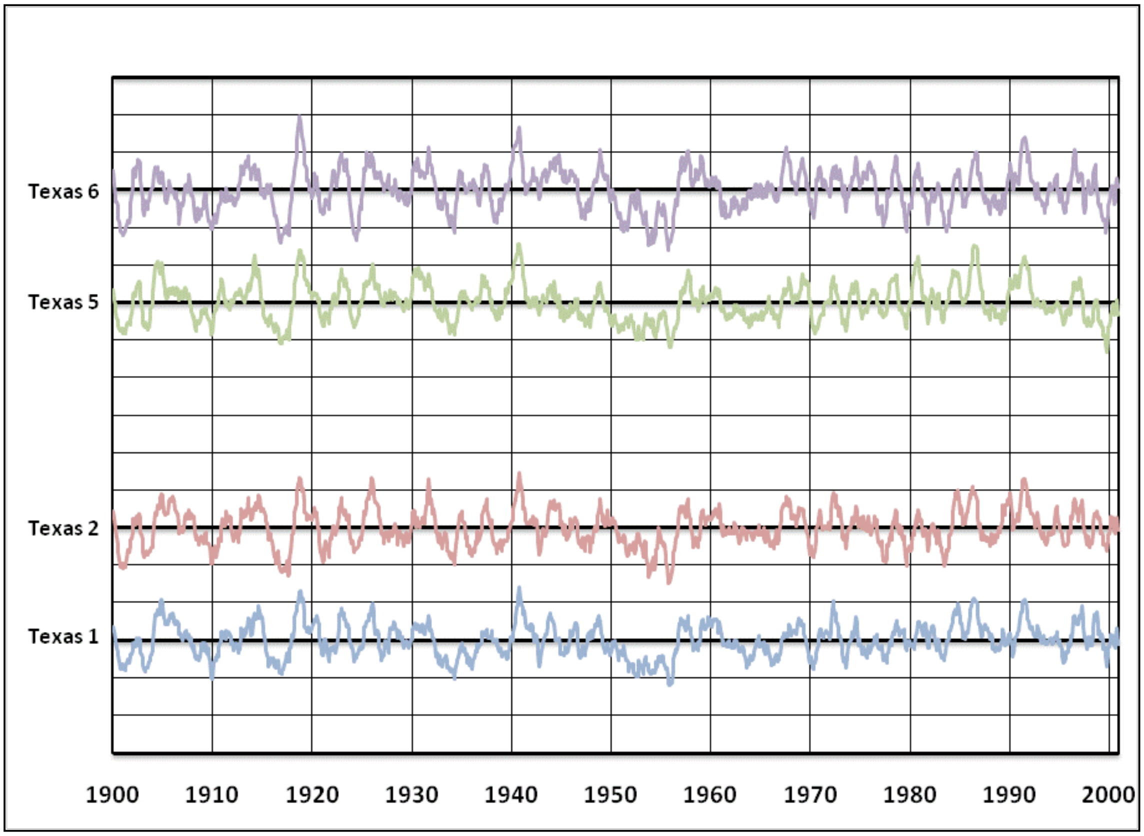


FIG 10.16. Time series of climate division averaged 12-month running precipitation MWZ values for the West Texas region using the CTY dataset. Each bold line represents a climate division averaged MWZ value of zero and each horizontal line an increment of two.

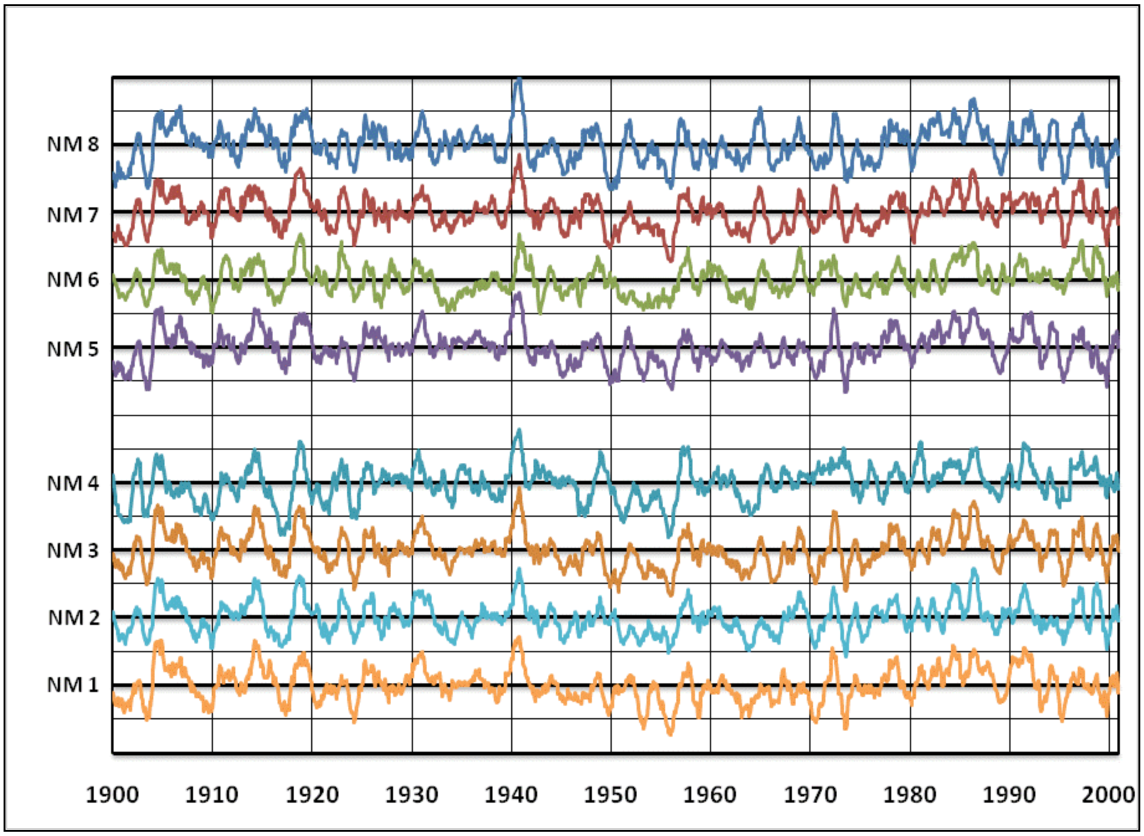
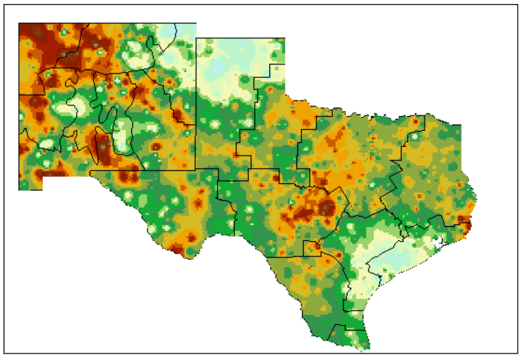
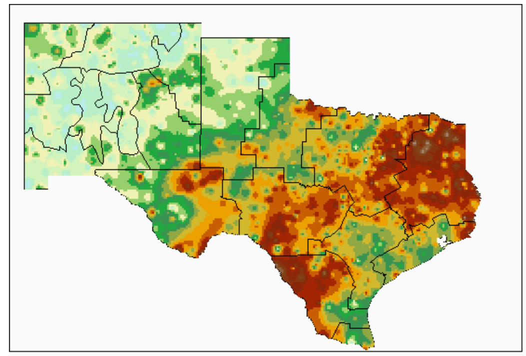


FIG 10.17. Time series of climate division averaged 12-month running precipitation MWZ values for the New Mexico region using the CTY dataset. Each bold line represents a climate division averaged MWZ value of zero and each horizontal line an increment of two.

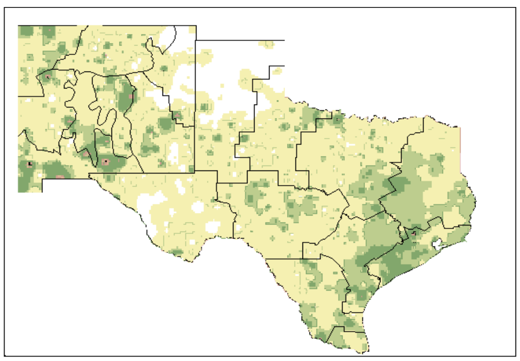
The last set of analyses looks at the 20th percentile (Figure 10.18) and the 2nd percentile (Figure 10.19) in a similar manner to the decadal maps showing drought percentages with good spatial density for the CqY dataset. The exception is that the century-long trend has been subtracted from each of the 12-month values; more plainly the CTY dataset is used. Again, the goal of this analysis is to judge drought on the assumption that the mean precipitation has not remained static over the past century and that the criteria for drought have changed along with a shift in the COOP station precipitation distributions. Figures 10.18 and 10.19 are sets of contour maps showing the spatial differences in drought across the ten decades of the 20th century.



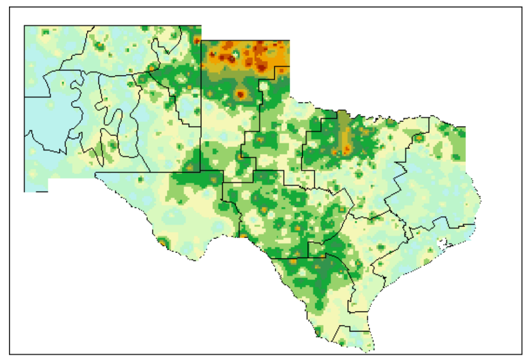
(a) 1900-1910



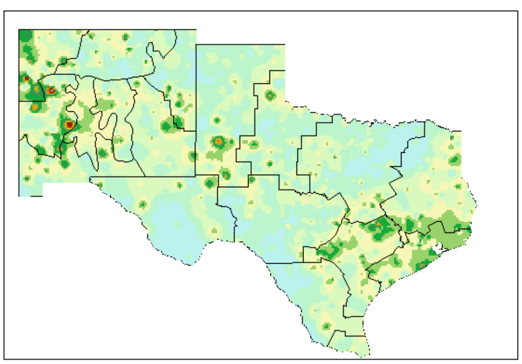
(b) 1910-1920



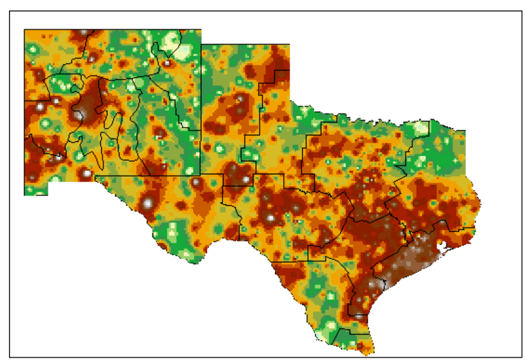
(c) 1920-1930



(d) 1930-1940



(e) 1940-1950



(f) 1950-1960

FIG 10.18. COOP station maps color-coded according to the percentage of months below the 20th percentile of its given distributions for the periods 1900-1910 (a), 1910-1920 (b), 1920-1930 (c), and 1930-1940 (d), 1940-1950 (a), 1950-1960 (b), 1960-1970 (c), and 1970-1980 (d), 1980-1990 (a), 1990-2000 (j) using the CTY dataset. The legend denotes the fractional percentage for the colors on the maps.

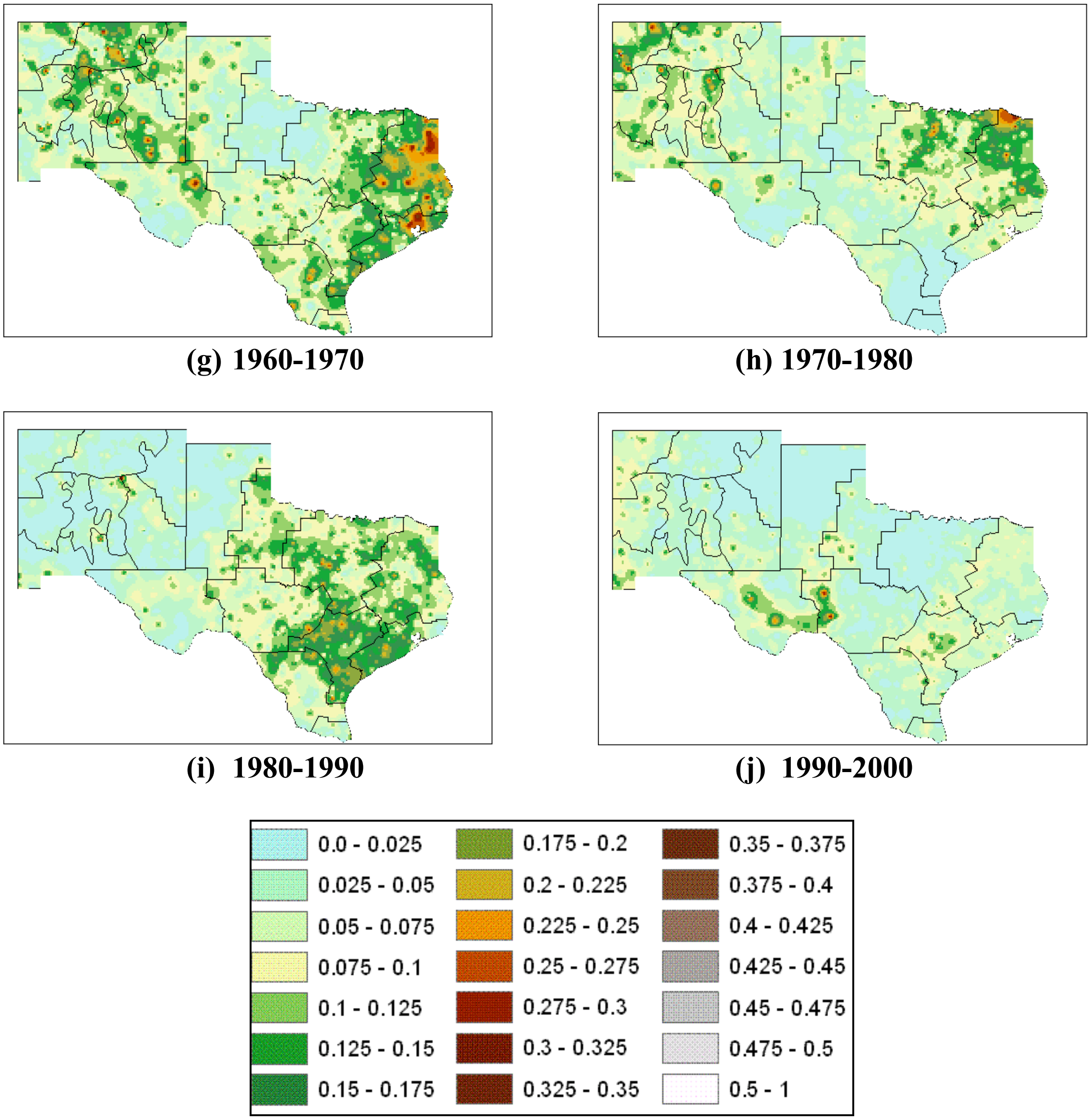
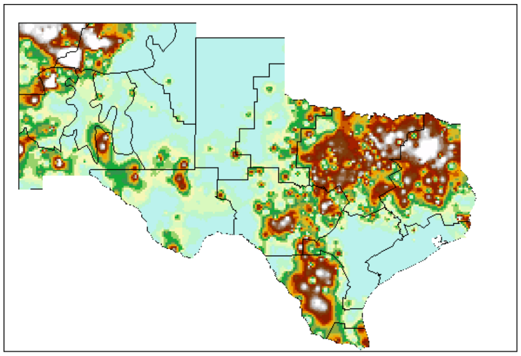
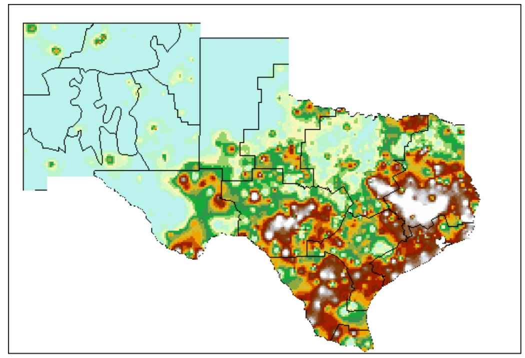


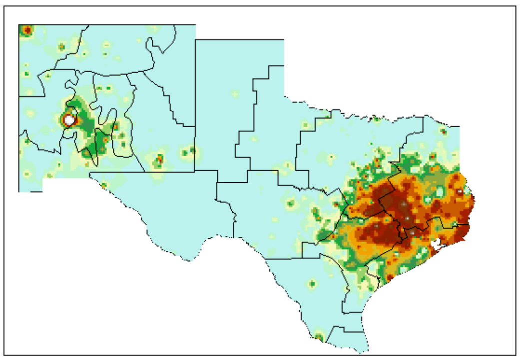
FIG 10.18. Continued.



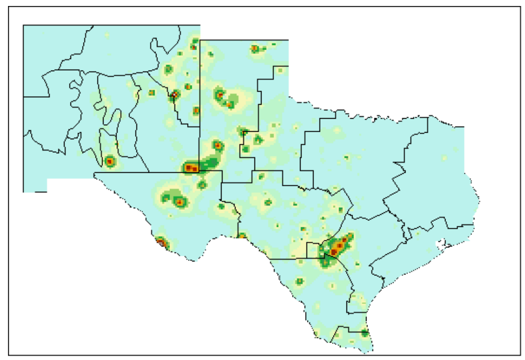
(a) 1900-1910



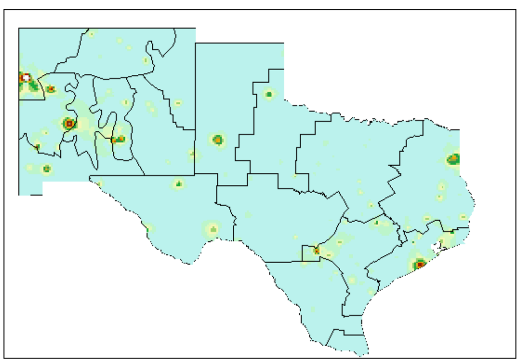
(b) 1910-1920



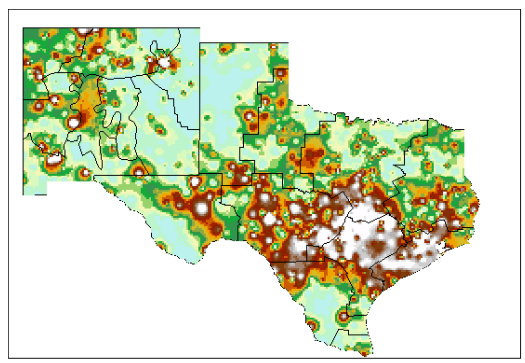
(c) 1920-1930



(d) 1930-1940



(e) 1940-1950



(f) 1950-1960

FIG 10.19. COOP station maps color-coded according to the percentage of months below the 2nd percentile of its given distributions for the periods 1900-1910 (a), 1910-1920 (b), 1920-1930 (c), and 1930-1940 (d), 1940-1950 (a), 1950-1960 (b), 1960-1970 (c), and 1970-1980 (d), 1980-1990 (a), 1990-2000 (j). using the CTY dataset. The legend denotes the fractional percentage for the colors on the maps.

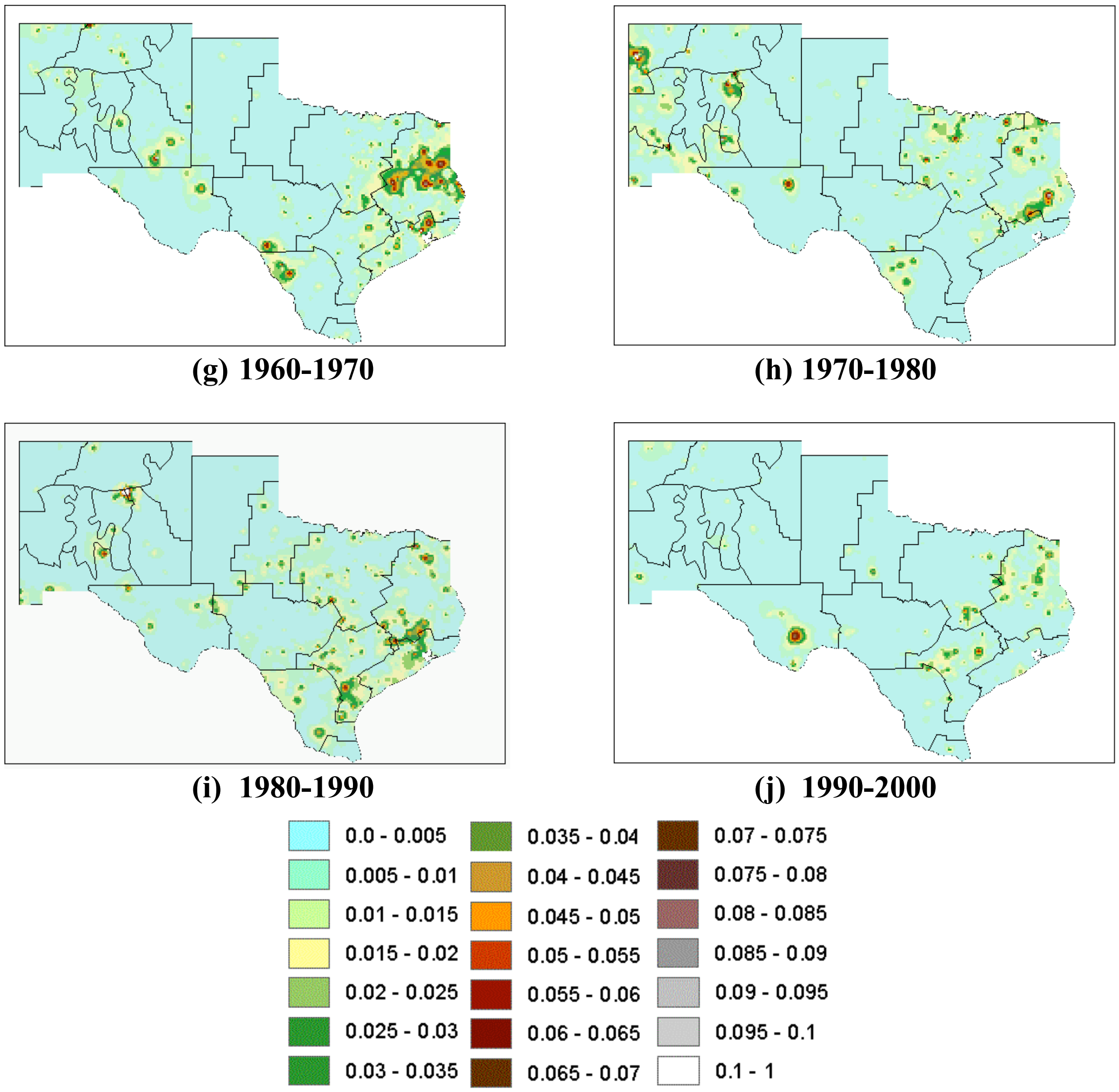


FIG 10.19. Continued.

11. CONCLUSIONS

The goal of the work done in the following study was to design a method to create a homogeneous datasets for the specific purpose of studying drought over the past century in Texas and New Mexico. This was a complicated process because there were a number of deterrants that can get in the way of transforming data at specific stations into time series data that are useful in terms of detecting climate trends. Of course, the only manner to study past climate trends that may signal drought or any other event is to have datasets that have disturbances not related to climate at a minimum.

Therefore, it was deemed vitally important to create a process by which to take the numerous COOP precipitation data available for study and transform these data into something useful for studying drought. Though temperature data were not explicitly analyzed in this specific study, steps were taken to create a process to homogenize these data as well. The IWSD interpolation scheme was chosen as the best candidate to accomplish these goals after a thorough investigation of several data interpolation schemes.

However, the Sun and Peterson (2005a) scheme was designed to investigate climate normals so some adjustment was needed to work with the monthly data that was the foundation of this study. Several tests were created in order to adjust the scheme to fit monthly data and this was possible due to the presence of the USHCN dataset created by NCDC. These highly quality-controlled data were used as the cornerstone of the interpolation process and any values created were indirectly related to this subset of stations.

The COOP dataset was a much more expansive dataset, but in order to maintain the integrity of the interpolation sheme designed in this study, not all of the available stations could be used. Therefore the creation of a thorough quality-control check of the COOP data was necessary to ensure that rogue values did not disrupt the interpolation process of the several analyses related to drought that followed. Also, the accuracy of the USHCN data values were not taken for granted so a fairly detailed metadata study was performed on the Texas and USHCN stations. The goal of this was to ensure that there was a subset of USHCN stations with a high degree of confidence in the homogeneity of the values.

The result of the interpolation process was the creation of several COOP and USHCN datasets, each of which had some aspect of the interpolation process tweaked. The result of each dataset were two distinct time series at each COOP and USHCN station, one containing the actual values and another containing the data from the interpolation process when actual values were not available.

The interpolated data are important not only because they are designed to reduce inhomogeneities but because they are more numerous. For the most part, the resultant interpolated data time series at each station were serially complete with values over the entirety of the 20th century. This was important because the vast majority of the COOP stations used in this study contained data records for less than half of the 20th century.

Further research into the distributions of monthly precipitation totals led to the approximation of these time series as containing data fitting a gamma distribution. Also, tests on the variance of the interpolated data found that the interpolateed process created distributions with artificially low variances. This was adjusted for with the creation of a third time series at each station that adjusted these interpolated variances to more realistic values.

Several of the datasets were investigated to determine the spatial characteristics of precipitation over the 20th century in Texas and New Mexico. A strength of the COOP dataset is its spatial density, but unfortunately these data are too numerous to show any specific statistical analyses so it was deemed necessary to group these data by climate division in order to show the trends and characteristics of these data. The analyses on precipitation trends showed that datasets agreed for the most part, but that the presence of the interpolated and variance-adjusted datasets compensated for some unrealistic values in the actual datasets.

The first of two datasets considered to have the most reliable data for analyzing past precipitation data were the USHCN dataset containing only the actual recorded data for stations containing century-long precipitation records. The second was the COOP datset containing the variance-adjusted data derived from the interpolation process in which only USHCN station with a high degree of confidence in their homogeneity were used.

The analyses created by the several datasets did not uncover any trends in climate or drought that were not known before the start of this project. However, the reduction of inhomogeneities in the datasets was the goal of this study because using data free from signals unrelated to natural climatic trends will be the key to understanding the behavior of drought. A clear understanding of past climatic trends preceding drought conditions, such as those in the 1950s, will enable those in charge to better allocate resources and prepare for droughts that are bound to happen in the future.

REFERENCES

Alexandersson, H., 1986: A homogeneity test applied to precipitation data. *J. Climatol.,* **6,** 661–675.

Alley, W. M., 1984: The Palmer Drought Severity Index: Limitations and assumptions. *J. Climate Appl. Meteor.,* **23,** 1100–1109.

Andreadis, K. M., E. A. Clark, A. W. Wood, A. F. Hamlet, and D. P. Lettenmaier, 2005: Twentieth-century drought in the conterminous United States. *J. Hydrometeor.,* **6,** 985–1001.

Barger, G. L., and H. C. S. Thom, 1949: Evaluation of drought hazard. *Agron. J.,* **41,** 519–526.

Byun, H.-R., and D. A. Wilhite, 1999: Objective quantification of drought severity and duration. *J. Climate,* **12,** 2747–2756.

Conrad, V., and C. Pollack, 1962: *Methods in Climatology*. Harvard University Press, 459 pp.

Davey, C.A., and R.A. Pielke Sr., 2005: Microclimate exposures of surface-based weather stations - implications for the assessment of long-term temperature trends. *Bull. Amer. Meteor. Soc*., **86**, 497–504.

Dickson, R. R., 1978: Weather and circulation from August 1978. *Mon. Wea. Rev.,* **106**, 1639-1645.

Easterling, D. R., T. C. Peterson, and T. R. Karl, 1996: On the development and use of homogenized climate datasets. *J. Climate,* **9,** 1429–1434.

Edwards, D. C., and T. B. McKee., 1997: Characteristics of 20th century drought in the United States at multiple time scales. Climatology Rep. 97–2, Department of Atmospheric Science, Colorado State University, Fort Collins, CO, May, 155 pp.

Eischeid, J. K., C. B. Baker, T. Karl, and H. F. Diaz, 1995: The quality control of long-term climatological data using objective data analysis. *J. Appl. Meteor.,* **34,** 2787–2795.

Friedman, D. G., 1957: The prediction of long-continuing drought in south and southwest Texas. Occasional Papers in Meteorology, No. 1, The Travelers Weather Research Center, Hartford, CT, 182 pp.

Gershunov, A., 1998: ENSO influence on intraseasonal extreme rainfall and temperature frequencies in the contiguous United States: Implications for long-range predictability. *J. Climate,* **11,** 3192–3203.

Gibbs, W. J., and J. V. Maher, 1967: Rainfall deciles as drought indicators. Bureau of Meteorology Bull. 48, Melbourne, Australia, 84 pp.

Gonzalez-Rouco, F., E. Zorita, H. Heyen, and F. Valero, 2000: Agreement between observed rainfall trends and climate change simulations in the southwest of Europe. *J. Climate,* **13,** 3057–3065.

Groisman, P. Ya., T.R. Karl, D.R. Easterling, R.W. Knight, P.F. Jamason, and Coauthors, 1999: Changes in the probability of heavy precipitation: Important indicators of climatic change. *Climatic Change,* **42,** 243–283.

Guttman, N.B., 2005: Spatial Regression as a technique for assessing the quality of data. *Extended Abstracts, 18th Conference on Probability and Statistics in the Atmospheric Sciences.* Asheville, NC, NOAA/NESDIS/NCDC, 1-2.

——, and R. G. Quayle, 1996: A historical perspective of U.S. climate divisions. *Bull. Amer. Meteor. Soc.,* **77,** 293–303.

Hecht, A.D., 1983: Drought in the Great Plains: History of societal response**.** *J. Appl. Meteor.*, **22**, 51–56.

Heim, R. R., Jr., 2002: A review of twentieth-century drought indices used in the United States. *Bull. Amer. Meteor. Soc.,* **83,** 1149–1165.

Henry, A.J., 1931: The Calendar Year as a unit of drought. *Mon. Wea. Rev.,* **59**, 150-153.

Jones, P. D., 1995: Land surface temperatures—Is the network good enough? *Climate Change,* **31,** 545–558.

Karl, T. R., and R. W. Knight, 1998: Secular trends of precipitation amount, frequency, and intensity in the United States. *Bull. Amer. Meteor. Soc.,* **79,** 231–241.

——, C. N. Williams Jr., F. T. Quinlan, and T. A. Boden, 1990: United States Historical Climatology Network (HCN) serial temperature and precipitation data. Environmental Science Division, Publication No. 3404, Carbon Dioxide Information and Analysis Center, Oak Ridge National Laboratory, Oak Ridge, TN, 389 pp.

Katz, R. W., and M. H. Glantz, 1986: Anatomy of a rainfall index. *Mon. Wea. Rev.,* **114,** 764–771.

Keim, B. D., A. M. Wilson, C. P. Wake, and T. G. Huntington, 2003: Are there spurious temperature trends in the United States Climate Division database? *Geophys. Res. Lett.,* **30,** 1404, doi:10.1029/2002GL016295.

Klein, W. H., and H. J. Bloom, 1987: Specification of monthly precipitation over the United States from the surrounding 700-mb height field. *Mon. Wea. Rev.,* **115,** 2118–3132.

Knappenberger, P., P. Michaels, and R. Davis, 2001: Nature of observed temperature changes across the United States during the 20th century. *Climate Res.,* **17,** 45–53.

Kunkel, K. E., D. R. Easterling, K. Redmond, and K. Hubbard, K. Andsager, M. Kruk, and M. Spinar, 2005: Quality control of pre-1948 cooperative observer network data. *J. Atmos. Oceanic Technol.,* **22,** 1691–1705.

Liles, C.A. 2003: Precipitation Variability in New Mexico - Some Causes and Consequences. Locally published for office use. Some excerpts have appeared in media.

Mann, H. B., and D. R. Whitney, 1947: On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Stat.,* **18,** 50–60.

Mauget, S.A., 2003: Multidecadal regime shifts in U.S. streamflow, precipitation, and temperature at the end of the twentieth century. *J. Climate*, **16**, 3905–3916.

Mendenhall, W., D. D. Wackerly, and R. L. Sheaffer, 1990: *Mathematical Statistics with Applications.* 4th ed. PWS-Kent, 788 pp.

Mo, K.C., and M. Chelliah, 2006: The Modified Palmer Drought Severity Index based on the NCEP North American Regional Reanalysis**.** *J. Appl. Meteor. Climatol.*, **45**, 1362–1375.

Momiyama, M. and M. Mitsudera, 1952. A stochastic study of climatology: Tendency of climatology. *Meteorology and Statistics,* **3***,* 171-177.

Mooley, D.A., 1972**.** Gamma distribution probability model for Asian summer monsoon monthly rainfall. *Mon. Wea. Rev.,* **101**, 160-176.

NCDC, 2002: Monthly precipitation probabilities and quintiles, 1971-2000. Climatography of the United States, no. 81, suppl. no. 1. Asheville, NC.

NCDC, 2005: Data documentation for Data Set 3206 (DSI-3200). National Climatic Data Center Tech. Doc. 3206, Asheville, NC, 18 pp.

NCDC, 2006: Data documentation for Data Set 3200 (DSI-3200). National Climatic Data Center Tech. Doc. 3200, Asheville, NC, 19 pp.

National Research Council, 2002: *Abrupt Climate Change: Inevitable Surprises.* National Academy Press, 244 pp.

Palmer, W. C., 1965: Meteorological drought. U.S. Weather Bureau Research Paper ***45***, 58 pp.

Peterson, T.C., D.R. Easterling, T.R. Karl, P.Y. Groisman, N. Nicholls, N. Plummer, S. Torok, I. Auer, R. Boehm, D. Gullett, L. Vincent, R. Heino, H. Tuomenvirta, O. Mestre, T. Szentimre, J. Salinger, E. Førland, I. Hanssen-Bauer, H. Alexandersson, P. Jones, D. Parker, 1998: Homogeneity adjustments of in situ atmospheric climate data: A review. *International Journal of Climatology*, **18**, 1493-1517.

Peterson, T.C. and R.S. Vose, 1997: An overview of the Global Historical Climatology Network. *Bull. Amer. Meteor. Soc.*, **78**, 2837-2849.

Quinlan, F.T., T.R. Karl, and C.N. Williams, Jr. 1987. United States Historical Climatology Network (HCN) Serial Temperature and Precipitation Data. NDP-019. Carbon Dioxide Information Analysis Center. Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, TN.

Sheppard, P. R., A.C. Comrie, G.D. Packin, K. Angersbach, and M.K. Hughes, 1999. The Climate of the Southwest. CL1-99. Institue for the Study of Planet Earth, University of Arizona, Tucson, AZ.

Stahle, D. W., and M. K. Cleaveland, 1988: Texas drought history reconstructed and analyzed from 1698 to 1980. *J. Climate,* **1**, 59–74.

Svoboda, M., D. LeComte, M. Hayes, R. Heim, K. Gleason, J. Angel, B. Rippey, R. Tinker, M. Palecki, D. Stooksbury, D. Miskus, and S. Stephens, 2002: The drought monitor. *Bull. Amer. Meteor. Soc.*, **83,**1181-1190.

Sun, B. and T.C. Peterson, 2005a: Estimating precipitation normals for USCRN stations, *J. Geophys. Res,* submitted.

Sun, B. and T.C. Peterson, 2005b: Estimating temperature normals for USCRN stations. *International Journal of Climatology,* **25***,* 1809-1817.

U.S. Department of Commerce, 1958: *Decadal Census of Weather Stations (Alabama through Wyoming less Alaska and Hawaii)*. U.S. Government Printing Office.

Wells, N., S. Goddard, and M.J. Hayes, 2004: A Self-Calibrating Palmer Drought Severity Index. *J. Climate,* **17***,* 2335–2351.

Wilks, D., 2006: *Statistical Methods in the Atmospheric Sciences*. Academic Press. 627 pp.

Wu, J., K. G. Hubbard, and J. K. You, 2005: Some concerns when using data from the Cooperative Weather Station Networks: A Nebraska case study, *J. Atmos. Oceanic Tech.,* **22***,* 592-602.

You, J, K.G. Hubbard, and S. Goddard, 2004: Comparison of estimates from spatial regression and inverse distance method, *J. Atmos. Oceanic Tech.,* submitted.

Young, K.W., 1993: Detecting and removing inhomogeneities from long-term monthly sea-level pressure time series. *J. Climate,* **6***,* 1205-1220.

APPENDIX A

TABLE A.1. The 221 USHCN stations used in this study. Stations in bold have data previous to 1918 and stations in italics were classified as high-quality based on the summary of metadata shown in Appendix B.

Station	Location	State	Latitude	Longitude	Elevation
020080	**AJO**	**AZ**	**32.37**	**-112.87**	**1800**
021026	**BUCKEYE**	**AZ**	**33.38**	**-112.58**	**890**
021614	**CHILDS**	**AZ**	**34.35**	**-111.70**	**2650**
023160	**FORT VALLEY**	**AZ**	**35.27**	**-111.73**	**7347**
024089	**HOLBROOK**	**AZ**	**34.90**	**-110.17**	**5070**
024849	**LEES FERRY**	**AZ**	**36.87**	**-111.60**	**3210**
025467	**MESA**	**AZ**	**33.42**	**-111.80**	**1235**
025512	**MIAMI**	**AZ**	**33.40**	**-110.88**	**3560**
026250	**PARKER 6NE**	**AZ**	**34.22**	**-114.22**	**410**
026796	**PRESCOTT**	**AZ**	**34.57**	**-112.43**	**5205**
027281	**ROOSEVELT 1WNW**	**AZ**	**33.67**	**-111.15**	**2205**
027370	**SACATON**	**AZ**	**33.07**	**-111.75**	**1285**
027435	**SAINT JOHNS**	**AZ**	**34.52**	**-109.38**	**5790**
027716	**SELIGMAN**	**AZ**	**35.32**	**-112.88**	**5250**
028619	**TOMBSTONE**	**AZ**	**31.70**	**-110.05**	**4610**
029271	**WHITERIVER 1SW**	**AZ**	**33.83**	**-109.97**	**5120**
029287	**WICKENBURG**	**AZ**	**33.98**	**-112.73**	**2095**
029359	**WILLIAMS**	**AZ**	**35.25**	**-112.18**	**6750**
029652	YUMA CITRUS	AZ	32.62	-114.65	191
030936	BRINKLEY	AR	34.88	-91.18	200
031596	**CONWAY**	**AR**	**35.08**	**-92.47**	**310**
031632	CORNING	AR	36.43	-90.58	300
032356	EUREKA SPRINGS 3WNW	AR	36.42	-93.78	1420
032930	GRAVETTE	AR	36.43	-94.45	1260
034572	MAMMOTH SPRING	AR	36.48	-91.53	650
034756	MENA	AR	34.57	-94.27	1130
035186	NEWPORT	AR	35.60	-91.28	228
035508	OZARK	AR	35.48	-93.82	390
035754	**PINE BLUFF**	**AR**	**34.22**	**-92.02**	**215**
035820	**POCAHONTAS 1**	**AR**	**36.27**	**-90.97**	**315**
035908	PRESCOTT	AR	33.80	-93.38	308
036928	SUBIACO	AR	35.30	-93.65	500
050848	BOULDER	CO	40.00	-105.27	5484
051294	CANON CITY	CO	38.42	-105.23	5330
051528	CHEESMAN	CO	39.22	-105.28	6880
051564	**CHEYENNE WELLS**	**CO**	**38.82**	**-102.35**	**4250**

TABLE A.1. Continued.

051741	**COLLBRAN**	**CO**	**39.25**	**-107.97**	**5980**
052184	DEL NORTE	CO	37.67	-106.35	7880
052281	**DILLON 1E**	**CO**	**39.63**	**-106.03**	**9065**
052432	**DURANGO**	**CO**	**37.28**	**-107.88**	**6600**
052446	**EADS 2S**	**CO**	**38.48**	**-102.78**	**4211**
053005	**FORT COLLINS**	**CO**	**40.58**	**-105.08**	**5004**
053038	FORT MORGAN 2S	CO	40.22	-103.80	4331
053146	FRUITA 1W	CO	39.17	-108.75	4480
053662	**GUNNISON 3SW**	**CO**	**38.53**	**-106.97**	**7640**
053951	HERMIT 7ESE	CO	37.77	-107.13	9000
054076	**HOLLY**	**CO**	**38.05**	**-102.12**	**3390**
054770	**LAMAR**	**CO**	**38.08**	**-102.62**	**3627**
054834	LAS ANIMAS	CO	38.07	-103.22	3890
055322	MANASSA	CO	37.17	-105.95	7690
055722	**MONTROSE #2**	**CO**	**38.48**	**-107.88**	**5785**
057167	**ROCKY FORD 2SE**	**CO**	**38.03**	**-103.70**	**4170**
057337	SAGUACHE	CO	38.08	-106.13	7692
057936	**STEAMBOAT SPRINGS**	**CO**	**40.50**	**-106.83**	**6840**
058204	**TELLURIDE**	**CO**	**37.95**	**-107.87**	**8672**
058429	TRINIDAD	CO	37.17	-104.48	6030
059243	**WRAY**	**CO**	**40.07**	**-102.23**	**3535**
160098	ALEXANDRIA	LA	31.32	-92.47	87
160205	AMITE	LA	30.70	-90.53	170
160537	BASTROP	LA	32.78	-91.90	150
160549	BATON ROUGE WSO AP	LA	30.53	-91.13	64
161411	CALHOUN RESEARCH STN	LA	32.52	-92.33	180
162151	**COVINGTON 4NNW**	**LA**	**30.53**	**-90.12**	**40**
162534	DONALDSONVILLE 4SW	LA	30.07	-91.03	30
163313	FRANKLIN 3NW	LA	29.82	-91.55	12
164407	HOUMA	LA	29.58	-90.73	15
164700	JENNINGS	LA	30.20	-92.67	25
165026	LAFAYETTE FCWOS	LA	30.20	-91.98	38
167344	PLAIN DEALING	LA	32.90	-93.68	290
168163	SAINT JOSEPH 3N	LA	31.95	-91.23	78
169806	WINNSBORO 5SSE	LA	32.10	-91.72	80
290692	*AZTEC RUINS NATL MONUMENT*	*NM*	*36.83*	*-108.00*	*5644*
290858	BELL RANCH	NM	35.53	-104.10	4500
291469	*CARLSBAD*	*NM*	*32.42*	*-104.23*	*3120*
291515	CARRIZOZO 1SW	NM	33.63	-105.88	5405
291664	CHAMA	NM	36.92	-106.58	7850

TABLE A.1. Continued.

291813	**CIMARRON 4SW**	**NM**	**36.47**	**-104.95**	**6540**
291887	**CLAYTON WSO AP**	**NM**	**36.45**	**-103.15**	**4970**
292848	ELEPHANT BUTTE DAM	NM	33.15	-107.18	4576
293265	***FORT BAYARD***	***NM***	***32.80***	***-108.15***	***6142***
293294	*FORT SUMNER*	*NM*	*34.47*	*-104.25*	*4025*
293368	GAGE 4ESE	NM	32.22	-108.02	4410
294369	**JEMEZ SPRINGS**	**NM**	**35.77**	**-106.68**	**6262**
294426	*JORNADA EXPERIMENTAL RANGE*	*NM*	*32.62*	*-106.73*	*4266*
295273	LUNA RS	NM	33.83	-108.93	7050
295960	*MOUNTAIN PARK*	*NM*	*32.95*	*-105.82*	*6780*
295965	MOUNTAINAIR	NM	34.52	-106.25	6520
296435	*OROGRANDE*	*NM*	*32.38*	*-106.10*	*4182*
297323	*RED RIVER*	*NM*	*36.70*	*-105.40*	*8676*
297867	**SAN JON**	**NM**	**35.12**	**-103.33**	**4230**
298107	*SANTA ROSA*	*NM*	*34.95*	*-104.68*	*4620*
298387	*SOCORRO*	*NM*	*34.08*	*-106.88*	*4585*
298501	SPRINGER	NM	36.37	-104.58	5922
299156	***TUCUMCARI 4NE***	***NM***	***35.20***	***-103.68***	***4086***
299165	TULAROSA	NM	33.08	-106.05	4430
340017	**ADA**	**OK**	**34.78**	**-96.68**	**1015**
340179	ALTUS IRRIGATION RES STN	OK	34.58	-99.33	1380
340256	ANTLERS	OK	34.25	-95.63	520
340292	**ARDMORE**	**OK**	**34.20**	**-97.15**	**840**
340548	BARTLESVILLE 2W	OK	36.75	-96.00	715
340593	BEAVER	OK	36.82	-100.53	2465
340908	BOISE CITY 2E	OK	36.73	-102.48	4145
341243	BUFFALO	OK	36.83	-99.62	1795
341504	CARNEGIE 2ENE	OK	35.12	-98.57	1290
341724	CHEROKEE	OK	36.77	-98.35	1180
341828	CLAREMORE 2ENE	OK	36.32	-95.58	588
342678	**DURANT-USDA**	**OK**	**34.02**	**-96.38**	**660**
342912	**ENID**	**OK**	**36.42**	**-97.87**	**1245**
342944	ERICK 4E	OK	35.20	-99.80	1985
343497	GEARY	OK	35.63	-98.32	1595
343628	GOODWELL RESEARCH STN	OK	36.60	-101.62	3310
343821	GUTHRIE	OK	35.88	-97.45	1030
343871	HAMMON 3SSW	OK	35.60	-99.40	1820
344055	HENNESSEY 4ESE	OK	36.10	-97.83	1150
344204	HOBART FAA AP	OK	35.00	-99.05	1552
344235	**HOLDENVILLE**	**OK**	**35.08**	**-96.40**	**860**

TABLE A.1. Continued.

344298	HOOKER	OK	36.87	-101.22	2995
344384	HUGO	OK	34.00	-95.52	570
344573	**JEFFERSON**	**OK**	**36.72**	**-97.80**	**1045**
344766	KENTON	OK	36.90	-102.97	4350
344861	**KINGFISHER 2SE**	**OK**	**35.85**	**-97.90**	**1100**
345063	LAWTON	OK	34.62	-98.45	1150
345509	MANGUM RESEARCH STATION	OK	34.83	-99.43	1520
345779	MEEKER 4W	OK	35.50	-96.98	925
345855	MIAMI	OK	36.88	-94.88	805
346130	MUSKOGEE	OK	35.77	-95.33	583
346139	MUTUAL	OK	36.23	-99.17	1865
346278	**NEWKIRK**	**OK**	**36.88**	**-97.05**	**1140**
346629	OKEENE	OK	36.12	-98.32	1210
346638	**OKEMAH**	**OK**	**35.43**	**-96.30**	**935**
346670	OKMULGEE WATER WORKS	OK	35.62	-96.02	647
346926	**PAULS VALLEY 4WSW**	**OK**	**34.73**	**-97.28**	**940**
346935	**PAWHUSKA**	**OK**	**36.67**	**-96.35**	**835**
347012	PERRY	OK	36.28	-97.30	1025
348501	**STILLWATER 2W**	**OK**	**36.12**	**-97.10**	**895**
348677	TAHLEQUAH	OK	35.93	-94.97	850
349395	**WAURIKA**	**OK**	**34.17**	**-98.00**	**875**
349422	WEATHERFORD	OK	35.52	-98.70	1642
349445	WEBBERS FALLS 5WSW	OK	35.48	-95.20	550
410120	**ALBANY**	**TX**	**32.73**	**-99.28**	**1420**
410144	**ALICE**	**TX**	**27.73**	**-98.07**	**201**
410174	***ALPINE***	***TX***	***30.37***	***-103.67***	***4480***
410493	**BALLINGER 2NW**	**TX**	**31.73**	**-99.98**	**1755**
410498	BALMORHEA	TX	30.98	-103.75	3220
410639	***BEEVILLE 5NE***	***TX***	***28.45***	***-97.70***	***255***
410832	**BLANCO**	**TX**	**30.10**	**-98.42**	**1370**
410902	**BOERNE**	**TX**	**29.80**	**-98.72**	**1422**
411048	**BRENHAM**	**TX**	**30.17**	**-96.40**	***313***
411138	BROWNWOOD	TX	31.72	-99.00	1385
411772	***CLARKSVILLE 2NE***	***TX***	***33.63***	***-95.03***	***435***
412015	*CORPUS CHRISTI WSO AP*	*TX*	*27.77*	*-97.50*	*41*
412019	***CORSICANA***	***TX***	***32.08***	***-96.47***	***425***
412121	**CROSBYTON**	**TX**	**33.65**	**-101.25**	**3010**
412266	***DANEVANG 1W***	***TX***	***29.05***	***-96.23***	***70***
412598	**DUBLIN**	**TX**	**32.10**	**-98.33**	**1502**
412679	***EAGLE PASS***	***TX***	***28.70***	***-100.48***	***805***

TABLE A.1. Continued.

412797	*EL PASO WSO AP*	*TX*	*31.80*	*-106.40*	*3918*
412906	***ENCINAL***	***TX***	***28.03***	***-99.42***	***590***
413063	***FALFURRIAS***	***TX***	***27.23***	***-98.13***	***120***
413183	***FLATONIA***	***TX***	***29.67***	***-97.12***	***520***
413280	FORT STOCKTON	TX	30.88	-102.87	2980
413734	***GREENVILLE***	***TX***	***33.15***	***-96.12***	***535***
413873	**HALLETTSVILLE 2N**	**TX**	**29.47**	**-96.95**	**275**
413992	**HASKELL**	**TX**	**33.17**	**-99.75**	**1600**
415018	**LAMPASAS**	**TX**	**31.05**	**-98.18**	**1024**
415196	**LIBERTY**	**TX**	**30.05**	**-94.80**	**35**
415272	**LLANO**	**TX**	**30.75**	**-98.68**	**1040**
415429	***LULING***	***TX***	***29.67***	***-97.65***	***398***
415618	***MARSHALL***	***TX***	***32.53***	***-94.35***	***352***
415707	*MCCAMEY*	*TX*	*31.13*	*-102.20*	*2450*
415869	***MEXIA***	***TX***	***31.68***	***-96.48***	***535***
415875	**MIAMI**	**TX**	**35.70**	**-100.63**	**2755**
416135	MULESHOE 1	TX	34.23	-102.73	3825
416276	**NEW BRAUNFELS**	**TX**	**29.73**	**-98.12**	**710**
416892	**PECOS**	**TX**	**31.42**	**-103.50**	**2610**
417079	**PLAINVIEW**	**TX**	**34.18**	**-101.70**	**3370**
417622	**RIO GRANDE CITY 3W**	**TX**	**26.38**	**-98.87**	**176**
417945	*SAN ANTONIO WSFO*	*TX*	*29.53*	*-98.47*	*788*
418201	SEMINOLE	TX	32.72	-102.67	3340
418433	**SNYDER**	**TX**	**32.72**	**-100.92**	**2335**
418692	**STRATFORD**	**TX**	**36.35**	**-102.08**	**3693**
418910	**TEMPLE**	**TX**	**31.08**	**-97.32**	**635**
419532	***WEATHERFORD***	***TX***	***32.77***	***-97.82***	***1065***
420086	ALTON	UT	37.43	-112.48	7040
420519	BEAVER	UT	38.30	-112.65	5940
420738	**BLANDING**	**UT**	**37.62**	**-109.48**	**6040**
420788	**BLUFF**	**UT**	**37.28**	**-109.55**	**4315**
421731	CORINNE	UT	41.55	-112.12	4220
422101	DESERET	UT	39.28	-112.65	4590
422418	ELBERTA	UT	39.95	-111.95	4690
422592	**ESCALANTE**	**UT**	**37.77**	**-111.60**	**5810**
422828	FILLMORE	UT	38.95	-112.32	5120
422996	FORT DUCHESNE	UT	40.28	-109.87	5050
423611	HANKSVILLE	UT	38.37	-110.72	4308
423809	HEBER	UT	40.50	-111.42	5630
423896	HIAWATHA	UT	39.48	-111.02	7280

TABLE A.1. Continued.

424508	KANAB	UT	37.05	-112.53	4950
424856	LAKETOWN	UT	41.82	-111.32	5980
425065	LEVAN	UT	39.57	-111.87	5300
425148	LOA	UT	38.40	-111.65	7070
425186	LOGAN USU	UT	41.75	-111.80	4790
425402	MANTI	UT	39.25	-111.63	5740
425733	MOAB	UT	38.58	-109.55	4043
425752	MODENA	UT	37.80	-113.92	5460
425826	MORGAN COMO SPRINGS	UT	41.03	-111.65	5080
426404	OGDEN PIONEER P H	UT	41.25	-111.95	4350
426601	PANGUITCH	UT	37.82	-112.43	6610
426686	PAROWAN POWER PLANT	UT	37.83	-112.83	6000
427260	RICHFIELD RADIO KSVC	UT	38.77	-112.08	5300
427318	RIVERDALE	UT	41.15	-112.00	4400
427516	SAINT GEORGE	UT	37.12	-113.57	2770
427714	SCIPIO	UT	39.25	-112.10	5300
427909	**SNAKE CREEK PH**	**UT**	**40.55**	**-111.50**	**6010**
428119	**SPANISH FORK PH**	**UT**	**40.08**	**-111.60**	**4720**
428705	THOMPSON	UT	38.97	-109.72	5100
428771	TOOELE	UT	40.53	-112.30	5070
428973	UTAH LAKE LEHI	UT	40.37	-111.90	4497
429111	VERNAL AP	UT	40.45	-109.52	5260
429382	WENDOVER AWOS	UT	40.73	-114.03	4237
429595	WOODRUFF	UT	41.53	-111.15	6315
429717	**ZION NATIONAL PARK**	**UT**	**37.22**	**-112.98**	**4050**

APPENDIX B

TABLE B.1. Number of station entries New Mexico and Texas USHCN stations and the locations that may have caused inhomogeneities in the data.

Station	Station type	Total number of station entries	Number of possibly entries causing inhomogeneities
290692	Rural	24	0
290858	Rural	7	1
291469	Small Town/Urban	20	0
291515	Rural/Small Town	17	3
291664	Mountainous	29	5
291813	Small Town	18	1
291887	Small Town	14	1
292848	Urban/Rural	14	1
293265	Rural	16	0
293294	Small Town	18	0
293368	Rural	10	1
294369	Mountainous	13	1
294426	Rural	18	0
295273	Rural	25	1
295960	Mountainous	13	0
295965	Small Town	26	5
296435	Rural	26	0
297323	Mountainous	13	0
297867	Rural	9	2
298107	Small Town/Urban	12	0
298387	Small Town	26	0
298501	Small Town	20	2
299156	Small Town	11	0
299165	Small Town/Urban	30	5
410120	Rural	17	1
410144	Urban	21	2
410174	Small Town	28	0
410493	Small Town	10	1
410498	Rural	12	1
410639	Rural	23	0
410832	Rural	11	1
410902	Small Town	24	3
411048	Urban/Small Town	16	0
411138	Urban/Small Town	14	1
411772	Rural/Small Town	20	0
412015	Urban	16	0
412019	Urban	10	0
412121	Rural	15	1
412266	Rural	14	0
412598	Small Town	10	1
412679	Small Town/Urban	13	0
412797	Urban	17	0
412906	Rural	18	0
413063	Small Town	17	0
413183	Rural	19	0
413280	Small Town/Urban	15	1
413734	Urban	16	0
413873	Rural	21	2

TABLE B.1. Continued.

413992	Rural/Small Town	16	1
415018	Small Town	19	1
415196	Rural/Small Town	15	1
415272	Small Town	16	2
415429	Rural/Small Town	19	0
415618	Urban	19	0
415707	Rural	17	0
415869	Small Town	11	0
415875	Rural	12	1
416135	Rural/Small Town	18	2
416276	Urban	21	1
416892	Small Town	10	0
417079	Small Town	18	1
417622	Small Town	16	1
417945	Urban	23	0
418201	Rural	19	3
418433	Small Town	21	2
418692	Rural	16	1
418910	Urban	14	2
419532	Small Town/Urban	17	0

APPENDIX C

This section is devoted to the interrogation of the 24 New Mexico and 44 Texas USHCN station metadata files to determine if the locations and elevations listed in those metadata files matched up with the description of the station location. The overall goal of the metadata study was to determine if inhomogeneities filtered into the USHCN precipitation analyses, since these stations form the baseline of our interpolation process. Davey and Pielke Sr. (2005) investigated the cause of spatial discrepancies in precipitation in the USHCN dataset and found that in the precipitation data, several stations suffered from observer bias.

Therefore it was deemed necessary to further investigate the observing practices at the USHCN stations by tracing the histories of the metadata files and noting anything unusual, most likely a suspicious location. This was a subjective analysis done with the use of an internet mapping tool called *Topozone*, which gives detailed political and topographical maps at most locations around the United States. The basic idea was to match up the location description denoted in the station metadata to the latitude and longitude also provided. An example of a station metadata file, abbreviated somewhat, is included in Table C.1 from USHCN station 412797 located in El Paso, TX.

Table C.1. USHCN Station 412797 abbreviated metadata file.

Dates of Interest		Lat	Lon	Station Movement	Elevation	Station Location	Station Description
11 06 1877	08 12 1880	31 47	106 30	999 999	3720	-	EL PASO/WBO
08 12 1880	11 01 1881	31 47	106 30	000 E	3720	-	EL PASO/WBO
11 01 1881	11 01 1882	31 47	106 30	000 W	3720	-	EL PASO/WBO
11 01 1882	04 01 1888	31 47	106 30	001 E	3720	-	EL PASO/WBO
04 01 1888	08 08 1894	31 47	106 30	001 NW	3720	-	EL PASO/WBO
08 08 1894	12 29 1907	31 47	106 30	000 NE	3720	-	EL PASO/WBO
12 29 1907	06 30 1925	31 47	106 30	002 NNE	3731	-	EL PASO/WBO
07 01 1925	04 28 1936	31 47	106 30	002 SSW	3720	-	EL PASO/WBO
04 28 1936	12 19 1942	31 47	106 30	003 E	3711	-	EL PASO/WBO
12 19 1942	05 07 1944	31 48	106 24	999 999	3920	57 ENE	EL PASO/WSO
05 08 1944	04 22 1959	31 48	106 24	999 999	3920	57 ENE	EL PASO/WSO
04 23 1959	08 31 1960	31 48	106 24	900 SW	3920	57 ENE	EL PASO/WSO
09 01 1960	04 01 1964	31 48	106 24	000 000	3918	57 ENE	EL PASO/WSO
04 01 1964	04 10 1978	31 48	106 24	003 SE	3918	57 ENE	EL PASO/WSO
04 10 1978	09 20 1978	31 48	106 24	000 000	3918	57 ENE	EL PASO/WSO
09 20 1978	11 13 1984	31 48	106 24	016 W	3918	57 ENE	EL PASO/WSO
11 13 1984	99 99 9999	31 48	106 24	000 000	3918	57 ENE	EL PASO/WSO

There is actually quite a bit more information in each line of a metadata file including the observer for each group of dates included and information about the instrumentation present at these times. All of the metadata files are useful in some aspect, but of most interest to this study are the dates of interest, latitude, longitude, direction and magnitude of station movement, elevation, station location, and station description columns denoted on Table C.1.

The history for USHCN station 41797 goes back to the year 1877 and is currently still in operation, with seventeen different periods of interest. The beginning of each period denotes a time in which one or more variables have changed, many times variables not noted in Table C.1. For instance stations entries may be due to instrumentation changes, changes in the person observing, or changes in the observing system. In addition to the metadata file, each station has an entry on the Multi-Network Metadata System kept by NCDC that contain useful remarks about station location.

Each entry in the metadata file contains a latitude and longitude listing with an elevation corresponding to its location. The second row on the metadata file contained in Table C.1 lists a station movement of “000 E,” which translates to moved less than a tenth of a mile to the east. Other magnitudes of station changes are shown in tenths of miles with the direction of movement. These station movement entries can be cues to whether a metadata file. Also of interest is the station location, which at almost every station is in relation to the post office. At the end of 1942, the station location of 412797 was 5.7 miles northeast of the post office and its name changed from “El Paso/WSO” to “El Paso/WBO.”

The detailed listing by *Topozone* allows for one to match up the coordinates to an elevation provided. Given that the coordinates provided in the metadata history of each station are not extremely precise, the elevation listed did not always exactly match the coordinates. For entries in which the elevation was reasonable for given coordinates, the station period was not deemed to be suspicious. However, if the elevation departure from the listed value was too much, the station entry was deemed suspicious.

Also of some use in this analysis was the use of the station descriptions in the Multi-Network Metadata System, which often gave a description of the environment surrounding the station. For instance, in the last entry for El Paso, the observing station was described as “located on a fairly level plain about five miles west of the Franklin Mountains.” If the description of the surrounding environment did not match the expectation of Topozone, the station entry was deemed suspicious. However, the suspicion in early entries was only warranted if the terrain description was not accurate, since buildings and land use can change over time.

An example of this was a station entry for USHCN Station 415272 in Llano, TX in the 1998. The description mentioned a move to a mile east-northeast of the post office to the sewer plant. However, *Topozone* did not verify this station location to be reasonable and the move was deemed suspicious. The location of the sewer plant in Llano was not in the general vicinity of the coordinates provided in the metadata. Often, the suspicious nature of move was based on a description in relation to a town's post office not matching the coordinates provided.

Another type of problem arose in USHCN station 413873 in Hallettsville, TX in which several elevations listed in the metadata did not correspond to the coordinates provided or the provided location in relation to the post office. Systematic errors in the listed metadata entries were another cause for suspicion in this study.

Through the use of *Topozone* the metadata histories of all the USHCN stations were examined thoroughly for potential biases that could possibly cause inhomogeneities in the climate record for that particular station. Stations deemed as "high quality" were deemed to have no such suspicious entries.

VITA

Name: Douglas Brent McRoberts

Address: Texas A&M Department of Atmospheric Sciences
3150 TAMU
College Station, TX 77843-3150

Email Address: mcrobert@tamu.edu

Education: B.S., Synoptic Meteorology, Purdue University, 2004
M.S., Atmospheric Sciences, Texas A&M University, 2008